

NATIONAL INSTITUTES OF HEALTH  
**NATIONAL LIBRARY OF MEDICINE**

PROGRAMS AND SERVICES  
FISCAL YEAR 1999

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES  
PUBLIC HEALTH SERVICE  
BETHESDA, MARYLAND

## National Library of Medicine Catalog in Publication

### National Library of Medicine Catalog in Publication

- Z  
675.M4  
U56an
- National Library of Medicine (U.S.)  
National Library of Medicine programs and services.--  
1977- . -- Bethesda, Md. : The Library, [1978-  
v. : ill., ports.  
Report covers fiscal year.  
Continues: National Library of Medicine (U.S.). Programs and services. Vols. for  
1977-78 issued as DHEW publication ; no. (NIH)  
78-256, etc.; for 1979-80 as NIH publication ; no. 80-256, etc.  
Vols. for 1981-available from the National Technical Information Service,  
Springfield, Va.  
ISSN 0163-4569 = National Library of Medicine programs and services.

1. Information Services - United States - periodicals 2. Libraries, Medical -  
United States - periodicals I. Title II. Series: DHEW publication ; no. 80-256, etc.

**DISCRIMINATION PROHIBITED:** Under provisions of applicable public laws enacted by Congress since 1964, no person in the United States shall, on the ground of race, color, national origin, sex, or handicap, be excluded from participation in, be denied the benefits of, or be subjected to discrimination under any program or activity receiving Federal financial assistance. In addition, Executive Order 11141 prohibits discrimination on the basis of age by contractors and subcontractors in the performance of Federal contracts. Therefore, the National Library of Medicine must be operated in compliance with these laws and executive order.

# CONTENTS

PREFACE .....	V
OFFICE OF HEALTH INFORMATION PROGRAMS DEVELOPMENT .....	1
NLM Long Range Plan .....	1
International Programs .....	1
Outreach Activities .....	3
LIBRARY OPERATIONS .....	5
Planning and Management .....	5
Collection Development and Management .....	6
Bibliographic Control .....	8
Information Products .....	11
Direct User Services .....	13
Outreach .....	14
Health Informatics Activities .....	17
SPECIALIZED INFORMATION SERVICES .....	22
Resource Building .....	22
Resource Access .....	23
AIDS Information Services .....	24
Outreach/User Support .....	25
LISTER HILL CENTER .....	26
Introduction .....	26
Knowledge Processing .....	27
Information Systems .....	29
Image Processing .....	34
Medical Informatics Training .....	38
Resource Support and Development .....	38
Engineering Laboratories .....	40
External Research Support .....	40
NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION .....	43
GenBank: The NIH Sequence Database .....	43
The Human Genome .....	44
PubMed .....	46
Other Specialized Databases and Tools .....	47
Database Access .....	48
Basic Research .....	49
User Support and Outreach .....	50
EXTRAMURAL PROGRAMS .....	52
Resource Grants .....	52
Training and Fellowships .....	52
Other Grants .....	54
Other Extramural Programs Activities .....	54
Grants Management Highlights .....	55

OFFICE OF COMPUTER AND COMMUNICATIONS SYSTEMS .....	58
Year 2000 Concerns .....	58
Development Branch .....	59
Systems Technology Branch.....	64
ADMINISTRATION.....	69
National Performance Review .....	69
Financial Resources .....	69
Personnel .....	70
NLM Diversity Council .....	75
INDEX .....	91
NLM ORGANIZATION CHART .....	inside back cover

#### APPENDIXES

1. Regional Medical Libraries.....	76
2. Board of Regents .....	77
3. Board of Scientific Counselors/LHC .....	78
4. Board of Scientific Counselors/NCBI.....	79
5. Biomedical Library Review Committee.....	80
6. Literature Selection Technical Review Committee.....	82
7. Extramural Programs Grants.....	83

#### TABLES

Table 1. Growth of Collections .....	19
Table 2. Acquisition Statistics.....	19
Table 3. Cataloging Statistics .....	20
Table 4. Bibliographic Services .....	20
Table 5. Circulation Statistics .....	20
Table 6. Online Searches—All Databases.....	21
Table 7. Reference and Customer Service.....	21
Table 8. History of Medicine Activities .....	21
Table 9. Extramural Grants and Contracts Program.....	57
Table 10. Financial Resources and Allocations.....	69
Table 11. Staff.....	74

## PREFACE

The National Library of Medicine lost the services of three outstanding long-time senior officials this year through retirement: Frances Humphrey Howard of the Division of Extramural Programs, Dr. Harold Schoolman of the Office of the Director, and Dr. Melvin Spann of the Division of Specialized Information Services. In different areas they have contributed much to the success of the Library's programs in recent decades. Retiring also, on September 30, 1999, was our legacy mainframe computer system, ELHILL. It has been replaced by systems with more modern retrieval capabilities, but it will still be missed by librarians and database searchers around the world. Among the notable gains of this year:

- MEDLINE*plus*, introduced in October 1998, has proved to be a popular Web site with links to consumer health information. Librarians from around the National Network of Libraries of Medicine contribute to its continually expanding content.
- In conjunction with the release of MEDLINE*plus*, NLM began a pilot project with some 40 public library systems in 10 states to see how consumer access to good health information could be improved. Network libraries were also invaluable in this project, working closely with the participating 200 local public libraries in nine states and the District of Columbia.
- LOCATOR*plus*, the new online catalog system that opens up NLM's collections of books, journals, audiovisuals, and manuscripts, became available in April 1999.
- Also in April, "Asthma: Breath of Life," a large-scale exhibit, was opened and continues to draw professional and lay visitors, including many from around the globe.
- In May, NLM made seven "Partners in Information Access for Public Health Professionals" awards to public health agencies and organizations around the U.S., bringing to 20 the total of such awards.
- MEDLINE accessed its 10-millionth citation in July. MEDLINE usage continues to climb steeply—from 100 million searches reported last year to the 191 million recorded in this report.
- In September 1999, the NLM received the prestigious "Hammer Award" from the Vice President's National Partnership for Reinventing Government.
- The NLM, on behalf of the NIH, is creating a database of extensive information about current clinical trials that will be introduced to the public in FY 2000.

I would like to express my appreciation to the staff of the Library for their continued dedication to the important work we do. It is they who earned the Hammer Award. My thanks go also to the advisors who serve on the Board of Regents and the Library's other advisory committees. The proof of their value is in the widespread acceptance and increasing contributions of NLM information services to the Nation's health.

---

Donald A.B. Lindberg, M.D.  
Director, NLM

# OFFICE OF HEALTH INFORMATION PROGRAMS DEVELOPMENT

*Elliot R. Siegel, Ph.D.*  
*Associate Director*

## **NLM Long Range Plan**

In January 1985, NLM's Board of Regents began to develop a 20-year Long Range Plan to guide the Library in using its human, physical, and financial resources to fulfill its mission. A broadly based process, involving the participation of librarians, health professionals, biomedical scientists, medical informaticians, computer scientists, and others whose interests were intertwined with the Library's, culminated in the adoption of a report by the Board of Regents published in 1987.

Planning reports prepared since then as supplements to the original Plan (also based on the work of broadly based panels of outside experts) contain recommendations on outreach to health professionals, electronic imaging, information services for toxicology and environmental health, the education and training of health science librarians, and, most recently, NLM's international programs. As a result of this ongoing planning effort, NLM has taken major strides forward. For example, the concept of a National Center for Biotechnology Information was born during a planning panel meeting.

Although the original plan and its supplements have served the Library well, a new Plan seemed timely for the new century. In order to develop the Long Range Plan 2000-2005, a document called The NLM Track Record was prepared as a synthesis and summary of past planning efforts and a statement of where we are today. The Track Record was sent out to over 250 past planning panel members and other advisors for comment as well as being posted on the public NLM web site. The comments received from well over 100 individuals were reviewed by the NLM Board of Regents, who asked staff to incorporate a number of priorities into a new draft Plan for the next three to five years. A broadly representative group of NLM advisors met on December 1, 1999, to review that draft. Based on their comments, and a final review by the Board of Regents, the NLM Long Range Plan 2000-2005 will be published in 2000 and will be available through the NLM web site.

## **International Programs**

### *Internet Connectivity at Malaria Research Sites in Africa*

“Communication with Atlanta and the rest of the world has improved tremendously. The importance of this cannot be overestimated. Contacts between scientists, between students and their professors and with our counterparts at CDC, contacts with companies and other institutes are a lot easier now. The system definitely is used. Literature searches, visits to malaria-related websites, but the main use is email.” (Erik Schoute, researcher, Centers for Disease Control, Kenya.)

NLM continues to lead the Communications Working Group (CWG) of the Multilateral Initiative on Malaria (MIM), which was begun in 1997. The objective is to support African scientists and the ability of malaria researchers to connect with one another and sources of information through full access to the Internet and the resources of the World Wide Web, as well as the creation of new collaborations and partnerships.

The initial meeting of the MIM CWG was held in January 1998 at the NLM/NIH in Bethesda, Maryland. In attendance were malaria research scientists, health information professionals, telecommunications experts and representatives of the major MIM funding agencies. In keeping with the underlying goal of supporting a broad spectrum of basic and operational malaria research needs, the researchers requested communications and connectivity capabilities sufficient to provide, at a minimum: robust and reliable e-mail, links to other research sites, access to full text journal articles, database searching, exchange of large files and mapping data, and timely access to electronic information resources worldwide.

In July 1999, Redwing Satellite Solutions Ltd. (based in the UK) and NLM's technical consultant Mark Bennett successfully installed VSAT ground stations at two malaria research sites in Kenya at Kisian (CDC funded) and Kilifi (Wellcome Trust funded). The 64kbs dedicated bandwidth purchased is shared by the two sites, with each paying \$3,000/month. These two sites join the Malaria Research and Training Center in Mali which has full Internet access via microwave technology, funded by NIAID and made operational in June 1998.

The NLM team is working to bring on three additional sites in the next six months: one in Nairobi (Walter Reed funded), and two in Ghana—in Accra

(Noguchi Institute) and in Navrongo (Navrongo Health Research Center). The Ghanaian sites, engaged in malaria vaccine testing, will be funded jointly by NIAID/NIH, the Naval Medical Research Center, and USAID. When these sites come on line, the overall bandwidth will increase to 128kbs and monthly charges will be reduced to \$2,100/month per site. Redwing has given NLM a group rate for the MIM sites and is treating the group of sites as one customer. The group arrangement has advantages. The rate per site would increase considerably if the site were to buy it on its own. In addition, the consortium approach allows for flexibility in adjusting bandwidth to fit the needs of the individual sites. NLM continues to support site visits and assessments, consultancies, evaluation and testing of the extant technology. Related issues of user training, in-country licensure of technology, and allowances for future technological advances (such as predicted worldwide availability of low-cost commercial satellite systems) are all part of an implementation plan developed by the MIM/CWG last year. The plan recommends immediate use be made of the affordable technologies now available to provide high-speed and reliable information and communication links in order to yield timely results in improving researchers' ability to do cooperative research and disseminate their results. Recommended technologies are VSAT, which uses a geostationary satellite and an earth station, and microwave, which uses radio waves. The latter is less expensive but is limited to line of sight transmission. The MIM sites that wish to operate a radio or VSAT link have to gain permission from the relevant in-country authority.

#### *International DOCLINE Libraries*

Under a new project of the NLM and the Medical Library of the University of Zimbabwe (UZML), malaria researchers in Southern, Central and East Africa can request the malaria-related documents and journal articles they need through the Medical Libraries at the University of Zimbabwe and South African Medical Research Council. Their requests will be filled either locally or at the NLM as quickly as possible, by Internet, e-mail, fax or airmail. This service will be provided free of charge as part of a pilot program funded by the Communications Working Group (CMG) of the Multilateral Initiative on Malaria at the NLM. Target countries for University of Zimbabwe Medical Library: Kenya, Malawi, Tanzania, Uganda, Zambia, and Zimbabwe. Target countries for South African Medical Research Council: Botswana, Swaziland, Angola, Namibia, Mozambique, and South

Africa. Under this new service, malaria researchers with Web access can search MEDLINE via the NLM's free PubMed service. During an online search, researchers can select journal references relating to their malaria research interests and request these references electronically using the PubMed's Loansome Doc feature. Note that some journals are linked full-text to PubMed references and can be obtained directly online (e.g., *British Medical Journal*).

#### *E-mail Applications*

For malaria researchers who would like to use e-mail as a document delivery method but who do not have document image application software (viewer), a free software application called DocView, developed by the Communications Engineering Branch of NLM, is available. DocView enables end users of libraries and information services to view, zoom, shrink, scroll, pan, rotate, and print bitmapped image documents received via e-mail. Documents can also be faxed or air mailed to researchers as the preferred method of delivery, especially for researchers without reliable e-mail connections or who prefer paper copies of documents.

**Special Malaria Web Site:** Malaria researchers with Web access are invited to access a new site of useful resources and links for malaria research that has been developed under the auspices of the Communications Working Group. The URL is for this rich resource is: <http://www.mimcom.net>.

#### *Global Internet Connectivity*

In 1999, NLM completed phase 2 of its end-to-end Internet connectivity testing and evaluation project. This project further explored the methods and metrics needed to better understand the quality of Internet performance from the end user perspective. To do this, NLM is using test methods that measure, for example, the size of the Internet transmission "pipe," the round-trip-time for sending packets of information to the destination and back, packet "loss," and response time. NLM collaborated with numerous domestic and international partners to test Internet pathways around the world. NLM monitored about 85 network paths to hosts (typically web servers) in 29 countries and in 14 time zones. Hosts were located in all major regions of the world: North, Central, and South America; Western and Eastern Europe; the Middle East; sub-Saharan Africa; and South, Southeast, and East Asia. The monitoring included locations in all G-7/8 countries. NLM conducted symmetrical (two-way) and time zone

testing with three partners as part of the G-7/8 Internet connectivity initiative: the Western Universities Research Consortium at the University of Calgary, Canada; OMNI (Online Medical Networked Information) at the University of Nottingham, England; and ForthNet, SA, in Crete, Greece.

NLM is now extending the Internet testing to include high bandwidth connections, including the vBNS (very high Bandwidth Network Services), with selected U.S. and international locations.

#### *International Visitors*

NLM receives many visitors of the international community from a variety of disciplines. Many of these visitors are responsible for medical, scientific, or technical information in their own countries. Visitors are briefed on relevant aspects of NLM operations and research. They also in most cases take tours of the NLM facilities. Among the visitors in FY 1999 were representatives from:

Australia, Belarus, Belgium, Brazil, Burma, Canada, Chile, China, Colombia, Croatia, Denmark, Egypt, England, Estonia, France, Georgia, Germany, Hong Kong, Hungary, India, Iran, Israel, Italy, Jamaica, Japan, Korea, Macedonia, Malawi, Malaysia, Mali, Mexico, Nigeria, Poland, Portugal, Romania, Russia, Saudi Arabia, Singapore, Slovenia, South Africa, Spain, Sri Lanka, Sweden, Switzerland, Taiwan, Tanzania, Thailand, Tunisia, Turkey, Turkmenistan, United Kingdom, Vietnam, Zambia, Zimbabwe

#### **Outreach Activities**

The NLM has a longstanding commitment to the effective dissemination and use of biomedical information within the health community. To help achieve this goal, NLM has, since 1989, collaborated with its National Network of Libraries of Medicine (NN/LM) to conduct outreach to health professionals and especially those in rural, minority, or other underserved communities. The objectives of NLM-sponsored outreach are to: 1) make health professionals aware of the information products and services NLM provides; 2) facilitate health professionals' access to and use of biomedical information; 3) provide training in the searching of electronic databases; 4) assist health professionals in developing new information-seeking behaviors; and 5) improve health care practices through the use of authoritative, up-to-date information.

Between 1990 and 1995, NLM supported close to 300 outreach projects that involved more than 500 institutions across the country. In 1996, NLM

published a five-year review of its outreach program and activities. The review concluded that NLM's outreach program has made significant progress overall. However, the review also recommended that the methodology for evaluating outreach be more fully developed.

The ability to conduct this five-year review was, to a great extent, facilitated by data collecting and reporting by the RMLs and NN/LM. This is an important first step in assessing the strategies that have been undertaken so far. However, to evaluate outreach approaches better and demonstrate the effectiveness of outreach projects in the future, several additional steps should be taken. NLM should develop benchmarks against which the RMLs and the NN/LM can measure the effectiveness of their outreach efforts. In addition, NLM and the RMLs should work together to develop further expertise in evaluation methodology. Evaluation components should be an integral part of all NLM-sponsored outreach.

The five-year review envisioned that strengthened outreach evaluation would help NLM and the NN/LM to more clearly discern lessons learned from past experience, better plan for future outreach activities, and design future outreach with a built-in evaluation component to the extent feasible. NLM determined that a logical next step in outreach would be to undertake a special project to develop a framework or model of outreach planning and evaluation. This approach is similar conceptually to that used by NLM with regard to telemedicine research, where the Institute of Medicine was commissioned to conduct a study and prepare a guide on evaluating the use of telecommunications in health care. NLM has required that its telemedicine contractors use the guide as part of a built-in evaluation component.

The working hypothesis is that the medical library community would benefit from knowledge of evaluation studies of outreach-like activities that have been carried out in related disciplines. NLM is especially interested in exploring related fields which have evaluated efforts directed toward minority populations because outreach to minority and other underserved populations is one of NLM's highest priorities, and, at the same time, an area in which success has been most difficult to achieve.

For this project on outreach evaluation, NLM selected the NN/LM's Pacific Northwest Regional Medical Library, located at the University of Washington Health Sciences Library. NLM decided that a sub-focus on Native Americans in the Pacific Northwest would be useful to better understand outreach to underserved populations and test the

outreach planning and evaluation field manual developed by the project. Initial project plans were presented at an outreach panel as part of the 1997 Medical Library Association meeting in Seattle. The project established an advisory panel consisting of representatives from academia, libraries, health professionals, and Native American groups. The panel has met twice. The field manual has gone through several drafts and panel reviews, and is now being reviewed and "field tested" by faculty, students, and practitioners in the library and information sciences.

A separate but related tribal connectivity project also administered by the Pacific Northwest

RML is providing support for new or enhanced Internet connections for 16 selected American Indian tribes and Alaska Native villages in the Pacific Northwest. Two of these sites are being used for pilot tests of the outreach planning and evaluation field manual.

Also, selected outreach planning and evaluation concepts were field tested as part of NLM's public library consumer health pilot project. Evaluation was built into this project from the outset. The evaluation results provided valuable information to NLM in making decisions about the future of outreach programs and projects involving public libraries.

# LIBRARY OPERATIONS

*Betsy L. Humphreys*  
*Associate Director*

NLM's Library Operations (LO) Division is responsible for basic services that provide access to the scholarly record of biomedicine and the health professions. LO selects, acquires, preserves, and organizes the world's largest collection of biomedical literature; produces authoritative indexing and cataloging records; builds and disseminates bibliographic, directory, and full-text databases; provides national back-up document delivery, reference, and research assistance; helps health professionals, researchers, librarians, and the general public to make effective use of NLM's services; and coordinates the 4,500 member National Network of Libraries of Medicine (NN/LM), which provides primary health sciences library services throughout the country. The basic information services provided by LO are an essential foundation for NLM's outreach programs to health professionals and the general public, as well as for its biotechnology, AIDS, and health services research information programs.

LO is the largest of NLM's divisions, employing a multidisciplinary staff of librarians, technical information specialists, subject experts, health professionals, historians, and technical and administrative support personnel. In addition to responsibility for basic services, the LO staff directs the National Information Center on Health Services Research and Health Care Technology (NICHSR); plays a key role in the development of NLM's Web information services; carries out an active program in the history of medicine; conducts research and evaluation studies related to the Library's programs and services; directs and participates in research in advanced information storage and retrieval; coordinates a post-graduate training program for medical librarians; and contributes to the development of Federal health data policy. LO staff members participate actively in Library-wide efforts to improve the quality of work life at NLM, including the Diversity Council and the expansion of the NLM Intranet.

## **Planning and Management**

LO's first priority is always to maintain and enhance the quality of basic services and operations while meeting increased user demand. In FY 1999, the LO staff was also focused on the following NLM-wide

priorities:

- replacing legacy automated systems used in internal operations and user services;
- planning for future space needs for the NLM collection and staff;
- expanding NLM services and programs to meet the health information needs of the general public; and
- defining workable strategies for acquiring, organizing, and providing permanent access to digital information.

LO also assisted in Year 2000 testing of critical NLM systems and helped the Planning Office prepare background documents in preparation for the development of a Library-wide strategic plan for 2000-2004.

LO continued to work closely with the Office of Computer and Communications Systems (OCCS) and other NLM program areas to manage the complex technical, organizational, and scheduling considerations involved in replacing essentially all of NLM's legacy systems and ending the Library's reliance on mainframe computers. In FY 1999, NLM implemented the Voyager Integrated Library System (ILS), an Oracle-based commercial product, which replaced multiple custom-developed systems supporting cataloging, acquisitions, basic serials processing, circulation, and online public access to NLM's catalog. The implementation of the new system required careful conversion and integration of literally millions of records from more than 15 different existing files, a major training program for hundreds of LO staff members, and ongoing adjustments to workflow and reallocation of staff to take best advantage of the new system's capabilities. In addition to implementing the ILS, LO and OCCS completed development of a new system for creation and maintenance of the MeSH thesaurus. Major progress was made on the development of replacement systems for online indexing and for DOCLINE and SERHOLD, which currently support automated document delivery request generation and routing in the National Network of Libraries of Medicine.

Work also began on the complex task of identifying and moving unique journal citations from specialized databases, e.g., HealthSTAR, AIDSLINE, HISTLINE, into the PubMed database and unique monograph citations into LOCATOR*plus*, the new Voyager online public access catalog. Direct online access to ELHILL ceased on September 30, 1999, 28 years after its first version launched nationwide online bibliographic retrieval.

In early 1999, LO staff completed a re-analysis of space needs for the NLM collection which indicated

that the Library would run out of space for the collection in 2003, if no special action were taken. By taking extraordinary measures, including moving some operations off-site, the Library can accommodate projected collection growth through the year 2008. Given a corresponding problem with space for NLM staff (due primarily to the establishment of the National Center for Biotechnology Information and other new programs after the Lister Hill Center Building was constructed), NLM commissioned special analyses of NLM's current space requirements and of the cost-benefits associated with various options (e.g., converting current space, renting or building offsite storage, building new contiguous storage) for future storage of the collection. LO staff members provided extensive information and assistance to the outside experts conducting the studies. In the meantime, LO was one of several NLM program areas that renovated some existing work areas to accommodate additional staff members. The main NLM reading room was renovated to include a raised floor for computer and other communications access, additional computer workstations, and a new Reference Desk. Raised flooring was also installed in the rotunda exhibition area to accommodate the computer and communications requirements of NLM's major exhibitions.

LO continued to play a major role in planning and implementing new and expanded services to meet the health information needs of the general public. LO staff members participated in the NLM-wide retreat on consumer health information in December 1998 and helped to draft the initial NLM policy for services to the general public that was approved by the Board of Regents in May 1999. NLM will focus on enhancing public access to electronic health information. LO directed the development of MEDLINE*plus*, NLM's new Web information service for the public; managed the pilot project to explore working with public libraries to improve the public's access to health information; and solicited special consumer outreach projects to be carried out through the National Network of Libraries of Medicine.

Like other research libraries, NLM must develop policies and procedures for acquiring, organizing, and ensuring access to the increasing amount of important information in digital form. NLM is interested in a wide range of electronic information: the Library's own electronic output, e.g., major databases such as MEDLINE and GenBank, technical reports, Web-based information describing NLM policies, programs, and services; scholarly publications, e.g., electronic journals and books; free Web-based information, such as that published by government

agencies, universities, and professional associations; and newly digitized information, such as retrospective indexes and full-text publications and unique unpublished manuscript collections.

In FY 1999, LO helped to define a general NLM strategy for making progress on the complex technical, organizational, and public policy issues affecting long-term access to digital information. This strategy involves using NLM products and services as test-beds for particular approaches, collaboration with other interested Federal and private sector groups, and developing realistic resource estimates for ensuring long-term access to digital biomedical information. The Library's current assessment is that ensuring permanent or perpetual "online" access is probably the most practical approach to guaranteeing preservation of many categories of digital information. In FY 1999, LO initiated or participated in several projects related to the management of electronic publications. These included: use of MEDLINE*plus* to test distributed approaches to describing health information on the Web; examination of NLM's current practices regarding the use of metadata in its own electronic publications, development of NLM requirements for licensing copyrighted electronic publications that are congruent with the Library's mission; and definition of policies and methods for ensuring permanent access to NLM's own electronic publications. LO is also assisting the National Center for Biotechnology Information in developing procedures for the planned PubMed Central electronic repository of research articles in the life sciences.

Additional information about activities related to NLM system reinvention, health information services for the public, and management of electronic publications appears elsewhere in this and other chapters.

## **Collection Development and Management**

NLM's comprehensive collection of biomedical literature provides essential support for many of the Library's services. LO's goal is to build and maintain a collection that serves the current and future needs of health professionals and researchers. To accomplish this, the LO staff develops and updates a formal literature selection policy, acquires and processes literature that meets its selection guidelines in all languages and formats, organizes and maintains the collection for efficient current use, and preserves the scholarly record of biomedicine for use by future generations. At the close of FY 1999, the NLM collection included 2.29 million volumes and 3.49

million other items, including audiovisuals, microforms, electronic publications, pictures, and manuscripts.

### *Selection*

The LO staff and the Library's book dealers and agents select literature for the NLM collection based on the guidelines in the Collection Development Manual of the National Library of Medicine. Historically this manual has undergone major revisions every five to ten years. In between major revisions, LO develops operational guidelines for emerging disciplines and new formats and for subjects that are posing particular selection problems.

As part of NLM's new initiative to improve access to health information to the general public, LO reviewed the current policy of acquiring only representative samples of print materials written for the general public for the NLM collection. The decision was made to retain the existing policy because larger holdings of print materials in NLM's physical collection will do little or nothing to improve public access to health information across the country. In selecting electronic information resources for its services for the general public, NLM will focus on science-based, nationally applicable resources that are available free or at low cost. During FY 1999, the Public Services Division developed and refined selection guidelines for MEDLINEplus, NLM's new Web information resource for the public. These guidelines are featured in the "About MEDLINEplus" section of that service.

LO staff members are participating actively in the NLM-wide effort to develop the Library's requirements for licensing copyrighted electronic publications. NLM's position as a national back-up document provider gives it relatively unique licensing requirements. The typical license is predicated on the ability to define a circumscribed group of primary users, which NLM does not have, and may limit use of materials for interlibrary loan, which is a basic NLM service.

### *Acquisitions*

LO received and processed 138,241 contemporary books, serial issues, audiovisuals, and electronic publications. (Table 2) Net totals of 33,155 volumes and 147,913 other items (e.g., microforms, audiovisuals, pictures, manuscripts) were added to the NLM collection during FY 1999. As expected, monograph acquisitions and serials processing were adversely affected by the transition to the new integrated library system. This was due to a

combination of unavoidable factors: downtime while data were converted from the old systems and loaded into the new one; the learning curve associated with transition to a totally new and quite different system; the need to initialize issue prediction patterns for all currently received serials; and, in the case of some serials processing tasks, the need to maintain the old and new systems in parallel until the reinvention of DOCLINE, SERHOLD, and the online indexing system is completed. By the fourth quarter, productivity for acquisitions and processing tasks had returned to normal levels. Despite the lower annual output, statistics for NLM's indexing and document delivery operations confirm that the Technical Services Division was successful in achieving its goal of maintaining good through-put for the highest priority materials.

The Library continued to add to its outstanding historical collections. Important early books acquired in FY 1999 included: *Gruntownij a Dokonaly Regiment...* (Prague, 1536) by Johannes Copp, a translation of a German work that is considered the first medical work of substance printed in Czech; Ambroise Pare's *Cinq Livres de Chirurgie* (Paris, 1572), treatises on bandaging, fractures, dislocations, toxicology, and gout designed for use by surgeons and considered Pare's chief work; Paracelsus' *Drey herrliche Schrifften...* (Basle, 1572), which explains the powers of the heart, lungs, brain, liver, kidneys, and spleen, and includes descriptions of remedies for treating ailments which afflicted these parts of the body; and *Hermetischer Probiere Stein* (Frankfurt, 1657) by Oswald Kroll, the German version of his *Basilica Chymica*, considered the standard work of iatrochemistry and one of the most important and influential works in the history of paracelsism. NLM also celebrated the return of *Recepta Varia*, a 12th century manuscript treatise on medicine produced in England, which had been missing from the NLM collection for nearly 50 years. The Library appreciates the assistance provided by the Wellcome Institute Library in London, the Louise Darling Biomedical Library at UCLA, and the Rootenberg family in effecting the return of this important early work.

Historical acquisitions in other formats included: a group of the sets of architectural plans for three Marine Hospitals (at Cincinnati, Galena, Illinois, and Wilmington, North Carolina) dating from 1850s; numerous additions to the poster collections, including six posters for "A Day without Art" in response to the AIDS epidemic from 1994 through 1998, four Belgian posters on the prevention of tuberculosis circa World War I, and fourteen posters on Christmas Seals from the 1920s through the 1950s; photographs of Native

American nurses from the Southwest acquired with the assistance of the Arizona Health Sciences Library; some 300 NIH-produced films from the 1940s, 1950s, and 1960s documenting research at the Institutes; and about 300 color slides of popular medical images from the 19th century photographed from books.

NLM received a deed of gift and the papers of NIH Nobel laureate Dr. Martin Rodbell, which will be digitized for the Profiles in Science database. Negotiations for acquisition of the papers of NIH Nobel laureates Dr. Julius Axelrod and Dr. Christian Anfinsen are continuing. Howard S. Lowensohn, Ph.D. donated the research files of Dr. Donald Gregg, former Head of the Department of Cardiology at Walter Reed who was named Federal Scientist of the Year in 1963 by President John F. Kennedy. Other additions to the manuscript collections and the NLM archives included documents from the American Board of Thoracic Surgery donated by Dr. Herbert Sloan; papers of Dr. Joseph McNinch, former Director of the Army Medical Library, donated by Eleanor McNinch; and the papers of Dr. Harold Schoolman, former NLM Deputy Director for Research and Education, that relate to his work on copyright issues over the past 30 years.

#### *Preservation and Collection Management*

To preserve NLM's archival collection and to keep it readily accessible for current use, LO binds, microfilms, conserves rare and unique items, maintains appropriate storage facilities and conditions for all types of library materials, and works to prevent and respond to emergencies that could damage these materials. LO distributes electronic data about what NLM has preserved to avoid duplication of effort by other libraries and provides preservation information useful to other health sciences libraries on NLM's Web site. NLM continues to monitor developments in preservation technology and to promote the use of more permanent media in new biomedical publications.

In FY 1999, LO bound 26,668 volumes, microfilmed 5,568 volumes, repaired 2,060 items in the onsite book repair and conservation laboratory, and conserved 174 items from the historical collections. One of the items conserved was *Recepta Varia*, the 12th century manuscript recovered by NLM in FY 1999 after being missing for about 40 years. Collection maintenance staff resolved or referred for repair 6,488 items with bibliographic or physical problems. A total of 4,283 old NLM microfilm reels were inspected; half were of unacceptable quality, and the Library will have to purchase or film replacements.

With the implementation of the Voyager

Integrated Library System, LO now uses barcodes to update individual item records when any physical item in the collection is circulated to the Reading Room or another location and again when the item is returned to the stacks. This is an additional workload, but it provides greatly enhanced inventory control. The reshelving contract was modified to include updating item records when items used in the Reading Room are returned to the collection.

Special collection and preservation management projects completed or initiated in FY 1999 included: shifting the modern manuscripts collection to create workspace in the History of Medicine Division for the Digital Manuscripts project staff; rehousing the pre-19th century thesis collection; reinforcing shelves in some areas of the stacks; inventorying and compressing the portion of the collection shelved by accession number; shifting serial issues published from 1985 to 1989 from the B1 level to compact shelving on the B3 level of NLM's stacks; inventorying the collection of 1850-1914 printed works; assessing the condition and collection management requirements of the historical film collection; and using invisible ink to record NLM ownership of approximately 600 rare books.

#### **Bibliographic Control**

To facilitate access to the biomedical literature, LO creates authoritative indexing and cataloging records for journal articles, books, films, pictures, manuscripts, and electronic media. LO updates and enhances the Medical Subject Headings (MeSH), the thesaurus used by NLM in its indexing and cataloging records and by many other institutions to describe the subject content of library materials and other types of information not acquired by NLM. MeSH is linked to other biomedical vocabularies and classifications within the Unified Medical Language System (UMLS) Metathesaurus, which is produced as a collaborative effort by the Lister Hill Center and LO. LO is also responsible for maintaining the National Library of Medicine Classification, a scheme for arranging physical library materials according to their subject matter used by health sciences libraries around the world.

#### *Thesaurus Development*

The 2000 edition of MeSH contains 19,768 main subject headings, 787 subheadings or qualifiers, 118 publication types, and 108,921 supplementary records for chemicals and other substances. Changes

made for the 2000 edition include 552 new descriptors, updated names for 130 descriptors, and more than 2,100 new cross-references. Special efforts were made to expand nervous system anatomy and nervous system disorders sections of the terminology. The nomenclature in NeuroNames, developed by the University of Washington's Regional Primate Information Research Center, was followed extensively in the expansion of the nervous system anatomy descriptors. (NeuroNames has been included in the UMLS Metathesaurus for several years.) The MeSH tree for symptoms and general pathology was revised and renamed "Pathological Conditions, Signs and Symptoms."

After a year of development and testing, OCCS and MeSH staff implemented a new Oracle-based MeSH maintenance system which replaces a previous system based on the Model 204 mainframe database management system. The new system will be used to produce the 2001 edition of MeSH. It employs a new concept-based structure, which reflects the organizing principle of the UMLS Metathesaurus. The new MeSH structure will support effective consolidation of other thesauri used in NLM databases such as POPLINE and BIOETHICSLINE within the MeSH system and will also allow the creation of a unified database that can facilitate the maintenance of translations of MeSH into other languages. To date, MeSH has been translated into more than fifteen different languages.

The majority of the content editing for the 2000 edition of the UMLS Metathesaurus was completed in FY 1999 under MeSH Section supervision. The MeSH Section worked with the Lister Hill Center to develop a three-day course in "Fundamentals of UMLS Metathesaurus Editing" for use in training new staff and contract editors of the Metathesaurus. New additions to the initial 2000 version of the Metathesaurus will include a clinical problem list terminology derived by Vanderbilt University Medical Center from actual problem statements entered in clinical records at Grady Memorial Hospital and Vanderbilt and selected drug concepts from First DataBank Master Drug Data Base and Multum MediSource Lexicon of FDA-approved prescription and OTC drugs.

### *Cataloging*

LO catalogs the biomedical literature acquired by NLM both to document what is available in the Library's collection and to provide cataloging records that can be used by other libraries to reduce the level of

effort they must expend to organize their own collections. Increasingly NLM is also cataloging or otherwise organizing information published on the World Wide Web—both to build new services, such as the MEDLINE*plus*, and to conduct tests that can contribute to a workable national strategy for organizing worthy health-related information available on the Web.

In FY 1999, the Technical Services Division cataloged 14,396 contemporary books, serials, non-print items (including some Web-based information sources), and cataloging-in-publication galley, using a combination of in-house staff and contractors. This was a significant decline from the previous year, due to expected system downtime, staff retraining, and changes in workflow associated with the implementation of the new Voyager Integrated Library System. Due to analogous declines in productivity in acquisitions, there was a slight decline in the backlog of uncataloged monographs. By the fourth quarter, cataloging production had returned to normal levels. The History of Medicine Division cataloged 58 monographs and 83 pictures. Enhancements to the commercial system underlying Images from the History of Medicine allowed HMD to begin to add records for additional images. The papers of the American Burn Association were processed and a finding guide for the collection was completed.

Cataloging Section staff contributed to the establishment of corporate name authorities for the Web information sources organized in MEDLINE*plus* under the direction of the Public Services Division and to the analysis of existing NLM practices for establishing metadata for NLM's electronic publications. Work was initiated to publish an updated print edition of the NLM Classification in FY 2000 and to make the Classification available on NLM's Web site.

LO staff worked to scale up production for Profiles in Science, a joint project of the Lister Hill Center and the History of Medicine Division that is digitizing and providing Web access to the papers of Nobel Prize winners and other prominent biomedical scientists. The project is designed to aid researchers and the general public and to develop procedures for digitizing manuscript collections. The creation of a Web site for each scientist involves digitizing documents (text, pictures, and videos), creating metadata or item-level catalog records for individual documents, authoring introductory biographical material, and creating an online exhibit designed for the public. Following the release of the initial site on Oswald T. Avery last fiscal year, the first installment of the papers of Dr. Joshua Lederberg became available in

March 1999. Substantial work was done on organizing, scanning, and creating metadata for the smaller collection of Dr. Martin Rodbell's papers in preparation for their release in early 2000.

### *Indexing*

LO indexes articles from more than 4,000 biomedical journals so that researchers, health professionals, and members of the public may locate articles on specific topics when they search MEDLINE or use any of its derivative products. In-house staff, contractors, and co-operating U.S. and international organizations perform the indexing under the supervision of the LO's Index Section. In addition to indexing newly published articles, LO also annotates existing MEDLINE records when the articles to which they refer have been retracted, corrected, or challenged in subsequently published notices or commentaries.

The Literature Selection Technical Review Committee (LSTRC) (Appendix 6), an NIH-chartered advisory group of outside experts, advises the Library about the journals that should be included in MEDLINE and *Index Medicus*. In FY 1999, the Committee reviewed 390 journals and rated 60 sufficiently highly for immediate inclusion in MEDLINE. NLM and the American Dental Association (ADA) met to discuss the coverage of dental journals in MEDLINE. A number of non-*Index Medicus* dental journals listed as indexed for MEDLINE have not been indexed for several years, due to resource and language limitations at the ADA, which is responsible for indexing them. The ADA agreed to review the dental journals and to identify the subset which should be indexed regularly. Given the ADA's plan to cease publication of the Index to Dental Literature after the 2000 issues, NLM will assume responsibility for ensuring that the subset of dental journals identified as high priority by the ADA is routinely indexed in the future. In conjunction with the development of the PubMed Central repository of electronic research reports in the life sciences proposed by the Director, NIH, NLM will also expand the coverage of life sciences literature in the MEDLINE subset of the PubMed database and in the larger PubMed database as well. The first step toward this goal was the decision to index from cover-to-cover some life sciences journals that have been indexed selectively for MEDLINE, as part of a general effort to reduce the number of selectively-indexed journals. LO is working with the National Center for Biotechnology Information to establish policies and procedures for expanding PubMed coverage in the life sciences.

In FY 1999, NLM added 434,000 indexed

citations to MEDLINE, which brought the number of MEDLINE citations to more than 10 million. Of the citations added this year, 43 percent were input via keyboarding, 33 percent by scanning and optical character recognition (OCR), and 23 percent were received in SGML-tagged electronic format direct from the publishers. However, by the fourth quarter of FY 1999, NLM had essentially achieved the goal of inputting one-third of MEDLINE citations by each of these methods. (This goal was established by the Director of NLM in January 1996 when 100 percent of MEDLINE citations were entered by keyboarding.) The Lister Hill Center continues to develop enhancements to the scanning/OCR operation to improve its efficiency. Journals that are more amenable to scanning are also more likely to move to direct submission of SGML-tagged data, so some gains in SGML input, which is the preferred of the three methods, may come at the expense of scanning/OCR input. One way or another, however, the amount of keyboarding is likely to continue to decline. At the close of FY 1999, NLM was receiving SGML-tagged data from 126 publishers for a total of 723 journals.

Irrespective of the initial data entry method, all MEDLINE citations are transferred to an online indexing system where indexers add subject headings and other data elements necessary to complete the citations. The current mainframe-based online indexing system is one of the many internal processing systems being replaced as part of NLM System Reinvention. During FY 1999, the Office of Computer and Communications Systems and LO made substantial progress on developing and testing the new indexing data creation and maintenance system, which is scheduled for implementation in FY 2000. The new system is Oracle-based, will use the Unicode character set, and will output citations in an XML format, both for PubMed and for licensees of NLM data.

LO continued to work with the Lister Hill Center on the Indexing Initiative research project, with the goal of identifying automated methods that can improve access to scientific literature without increasing the level of human resources required. LO conducted initial experiments to assess the feasibility of automated indexing of gene names in journals known to publish articles related to yeast genes. The results were promising, but, since yeast is in some ways an ideal case, follow-on experiments involving other organisms are needed.

### **Information Products**

NLM produces online databases, other

electronic products, and print publications incorporating its authoritative indexing, cataloging, and thesaurus data. LO collaborates with other NLM Divisions to produce and distribute some of the world's most heavily used medical information resources.

#### *Databases and Web Information Resources*

Use of the MEDLINE database via PubMed and Internet Grateful Med (which interfaces with the PubMed database and its Entrez retrieval system) now exceeds 18 million searches per month, a 60 percent increase from the previous year. Approximately 80,000 unique "IP" addresses access PubMed each day. As planned, use of the ELHILL retrieval system dropped 81 percent to about 3 million searches for the entire year. LO assisted the National Center for Biotechnology Information in the development and testing of a new release of PubMed that includes many features requested by librarian searchers. A beta version of this release became available for user testing and feedback late in FY 1999. LO staff also worked with the Lister Hill Center on the design and development of a new NLM Gateway, with the goal of providing a simplified interface which can search and return relevant results from multiple NLM systems. As an interim step in this direction for NLM's health services research databases, NICHSR and the Lister Hill Center implemented an HSR Search capability which provides a single point of entry to search the HealthSTAR bibliographic database, the HSRProj research projects database, the HSTAT full-text database, the prototype HSR Tools database, and health services research organizations in DIRLINE. Both PubMed and HSTAT also have reciprocal links with the new National Guidelines Clearinghouse sponsored by the Agency for Health Care Policy and Research, the American Medical Association, and the American Association of Health Plans.

In a major milestone for NLM System Reinvention, direct online access to ELHILL ceased on September 30, 1999. Twenty-eight years earlier the first version of ELHILL ushered in the era of online searching via commercial telecommunication networks. ELHILL evolved to a blindingly fast, highly reliable, sophisticated command language search system that could perform miracles of precise retrieval in the hands of expert searchers. However, ELHILL was not ideally suited to interact with personal computer interface programs, as annoying technical glitches in successive versions of Grateful Med illustrated. It could not easily support statistical matching, ranking, and clustering retrieval techniques that are especially useful to people

without special search training. Nor could it be used to embed pre-formulated MEDLINE searches into Web pages, as is easily accomplished with PubMed. ELHILL also required programming expertise of a type that has essentially disappeared from the marketplace. Despite technical limitations that increasingly prevented NLM from developing new services and retrieval capabilities, ELHILL's speed, reliability, and sophisticated search capabilities proved very difficult to replace. If "legacy system" can be defined as "a system that works," ELHILL was the quintessential legacy system.

In another major System Reinvention achievement based on the implementation of the Voyager integrated library system, NLM released a new online public access catalog, LOCATOR*plus*, in April 1999. The new Web-based catalog combines information previously found in the separate CATLINE, AVLINE, and SERLINE files; indicates the current availability of specific items; presents titles retrieved in rank order; and supports direct Web links to the full-text of titles described in the database. In FY 1999, the first step in moving the citations from specialized NLM databases into the PubMed database (for the journal citations) and LOCATOR*plus* (for monograph and chapter citations) was accomplished when journal citations from HealthSTAR were added to PubMed in June 1999. LO and OCCS expect to complete this process for HealthSTAR, AIDSLINE, HISTLINE, SPACELINE, BIOETHICSLINE, and POPLINE in FY 2000. Early in 1999, LO and the Lister Hill Center developed and released a Web-based MeSH browser that provides access to the full MeSH records needed by indexers and catalogers, replacing the ELHILL MeSH file.

Citations from the 1960-63 volumes of *Index Medicus* were loaded into the OLDMEDLINE file, which now includes nearly 800,000 records. Contracts were awarded to keyboard 1957-1959 Current List data. Lakota Technologies, Inc., a Native American Organization, will keyboard the January-June 1957 data. A contract was awarded to MCP/Hahnemann University for digitization of the *Index Catalogue of the Library of the Surgeon General*, an important source for historians of medicine and science. Work was completed on rescanning at higher resolution all pictures in Images from the History of Medicine. The higher quality images, complete with new "digital watermark," will be viewable on NLM's Web site in FY 2000.

On October 22, 1998, NLM released MEDLINE*plus*, a new Web-based health information source for the general public. By the end of FY 1999, users were accessing this new service at the rate of

about 1 million pages per month. MEDLINE<sub>plus</sub> indexes individual Web documents produced by the National Institutes of Health and other authoritative sources and also contains pre-formulated MEDLINE searches designed to retrieve references to articles likely to be suitable for a non-specialist readers. During FY 1999, the number of topics covered by MEDLINE<sub>plus</sub> increased from 22 to about 225. LO contracted with health sciences libraries at the University of Cincinnati and Indiana University and with several individual health sciences librarians for assistance in selecting and describing high quality Web pages for inclusion in the site. Both inhouse and remote contract staff use the same maintenance system developed by OCCS using a combination of Oracle relational database, Cold Fusion, Java scripts, and Web technology. This application supports efficient, distributed creation, review, maintenance, and monitoring of links to specific Web pages.

The Library released a substantially redesigned version of MEDLINE<sub>plus</sub> on September 22, 1999. The redesign was based on recommendations and data obtained from a variety of sources. Individual users of the site used the MEDLINE<sub>plus</sub> "comment" feature to provide feedback. Librarians participating in the pilot project to provide health information in about 40 public library systems provided input in regularly scheduled conference calls, reports, and site visits. NLM convened an Ad Hoc Advisory Group on Consumer Health Information in May 1999 to review MEDLINE<sub>plus</sub> (with a focus on the topic of diabetes) and to provide recommendations on the Library's initiative to improve the public's access to health information. Chaired by NLM Director Emeritus, Dr. Martin M. Cummings, the ad hoc group included health professionals, health educators, librarians, and consumer representatives. LO also commissioned reviews by experts in human-computer interface design and arranged for a formal usability test in which members of the public were videotaped as they attempted to use MEDLINE<sub>plus</sub> to answer specific questions. Staff members also studied Web logs to identify user navigation problems.

#### *Machine-Readable Data*

NLM leases its databases in machine-readable form to promote the broadest possible use of its authoritative bibliographic and thesaurus data. Commercial companies, international MEDLARS centers, universities, and other interested organizations then make NLM data available online or in CD-ROM products or use them to improve the functionality of a variety of medical information systems. In FY 1999, the

Library distributed MEDLINE and other bibliographic databases to 53 organizations. Hundreds of organizations and individuals obtained MeSH data via file transfer protocol from the Internet. More than 1,100 organizations and individuals license the UMLS Knowledge Sources and associated programs. The data are available on CD-ROM, via ftp, or through the applications programming interface or interactive use of the UMLS Knowledge Source Server developed and maintained by the Lister Hill Center.

The reinvention of the Library's internal systems inevitably has an impact on the distribution of its machine-readable data. Following the implementation of the new integrated library system, OCCS completed new MARC conversion programs for NLM's cataloging and authority records. NLM now distributes MARC-formatted cataloging records for books, serials, and non-print media in a single CATFILE. The Library also produces a separate MARC-formatted SERFILE that contains only the serial records, with some additional data elements of interest to those who previously leased the ELHILL SERLINE database. With the transition away from ELHILL and the implementation of the new indexing data creation and maintenance system, NLM will change to a new XML-format for distribution of MEDLINE and other journal citations in 2000. LO and OCCS staff defined and tested the new format, which will use the Unicode character set. When the new XML format is implemented, NLM will distribute weekly MEDLINE updates via ftp and the full files on DLT tape.

#### *Print and Electronic Publications*

NLM publishes some of its authoritative data in printed publications, including *Index Medicus* and several MeSH publications. The Library continues to review and modify or eliminate specific print publications that have outlived their usefulness, given increasing user access to more flexible forms of NLM data.

As NLM replaces its legacy systems, the programs that generate any publications the Library intends to continue must also be replaced. During FY 1999, LO assisted OCCS in defining and testing new methods for generating *Index Medicus* and the *List of Journals Indexed in Index Medicus*. LO has decided not to print the 2000 *List of Serials Indexed for Online Users*, but it will be available in pdf format for local printing elsewhere. For many years, NLM has produced camera ready copy for the *Index to Dental Literature*, which is published by the American Dental Association,

and for the *Hospital and Health Administration Index*, which is published by the American Hospital Association. The American Dental Association has informed NLM that it intends to cease publication of the *Index to Dental Literature* after the 2000 edition. The American Hospital Association will cease publication of the *Hospital and Health Administration Index* after the 1999 edition.

The NLM World Wide Web site continues to grow as the primary vehicle for distributing a wide range of NLM publications from fact sheets to extensive reports. There were 2.03 million hits to NLM publication pages in FY 1999, a 31 percent increase from the previous year and about 10 percent of the overall traffic on the NLM Web site. Among the most popular publications are issues in the Current Bibliographies in Medicine series, some of which are also available in print. Each bibliography addresses a topic of current interest that may be difficult to search in NLM's databases or be spread across the literature of multiple disciplines. The bibliographies cover topics that are important to specific programs of the NLM, NIH, or other Federal agencies. They are often produced as background for NIH Consensus Development Conferences or other special meetings. LO staff members collaborate with outside experts to produce each bibliography. FY 1999 additions to the series included: Ethical Issues in Research Involving Human Participants, Trimethylaminuria (Fish-Malodor Syndrome) and the Flavin Monooxygenases, and Multicultural Aspects of Breast Cancer Etiology.

During FY 1999, LO staff members led a major redesign of the NLM Web site to help site visitors find health information more easily, to create a simpler, cleaner, more consistent look across NLM pages, and to organize NLM programs and services into more understandable categories. Like the MEDLINEplus redesign, this effort assisted by the comments of experts in interface design and by examinations of Web logs to identify common problems experienced by visitors to the site. The new site design debuted on July 22, 1999. User feedback was very positive, and follow-up analysis of Web logs indicates that users are getting to sources of health information more effectively.

### **Direct User Services**

In addition to its electronic and printed products, LO provides document delivery, reference, and customer service to remote users as a national back-up to services available from other health sciences libraries and information suppliers. LO also serves a

large onsite clientele in the NLM reading rooms.

### *Document Delivery*

LO provides copies of documents in the NLM collection to other U.S. and international libraries to fill requests from health professionals, researchers, and other interested people which cannot be readily supplied by other members of the National Network of Libraries of Medicine, health sciences libraries in other countries, or other document suppliers. LO also retrieves documents from the Library's closed stacks for use by onsite patrons.

In FY 1999, LO received a total of 751,732 requests for documents, an 8 percent increase from the previous year. NLM's onsite users requested 355,216 documents, an 11 percent increase from FY 1998. In April 1999, LO and OCCS implemented the closed stack circulation module of the Voyager Integrated Library System, which was developed at NLM's request. This module requires that each item sent to an NLM Reading Room be explicitly charged out to a specific patron and discharged when the item is returned to the stacks, providing much better tracking of items in the NLM collection and the ability to analyze onsite use. While testing the new system, LO staff replaced time-consuming keyboard entry of charge-out commands with pre-printed barcode commands entered by electronic barcode readers, thus avoiding an increase in the time required to provide materials to onsite users. A "link to local holdings" was established for most of the SilverPlatter WebsPIRS subscription databases that LO makes available to onsite patrons and staff. The link connects an onsite user from the SilverPlatter bibliographic record to NLM's LOCATORplus catalog on the Internet.

The number of interlibrary loan requests increased 6 percent to 396,516. Seventy-six percent were filled, up from 74 percent in FY 1998. Twenty-eight percent of the filled requests were filled electronically. Throughput rates were outstanding: 80 percent of requests were processed within 8 hours of receipt, 90 percent within 24 hours, and 95 percent within 48 hours. The FY 1998 implementation of the Relais automated document request tracking and fulfillment system has both speeded throughput and reduced transaction costs. A new version of Relais installed in the fourth quarter of FY 1999 provided additional workflow enhancements.

Relais enhances document request processing at NLM, but the DOCLINE system continues to route ILL requests among U.S. health sciences libraries and an increasing number of international libraries. The

3,054 DOCLINE libraries entered 3.02 million requests into the system, an increase of 4 percent from FY 1998; 12.5 percent of the requests were routed to NLM. The overall DOCLINE fill rate held steady at 94 percent. By September 1999, 67 percent of system use was via the Internet, up from 53 percent in October 1998. Use of the Loansome Doc feature, which allows individuals searching via PubMed or Internet Grateful Med to route document requests to a health sciences library which agrees to serve them, increased 10 percent to 448,752 requests. Many international PubMed and Internet Grateful Med searchers are requesting Loansome Doc service. In an effort to serve these users, LO has enlisted a number of U.S. and international libraries which are willing to supply documents to individuals outside the U.S. The libraries and the countries they will serve are listed on the NLM Web site.

#### *Reference and Customer Service*

LO provides reference and research assistance to onsite and remote users as a backup to services available from other health sciences libraries. LO also has primary responsibility for responding to inquiries from those seeking information about NLM products and services or assistance in using these services. LO received a total of 111,279 inquiries from onsite and offsite users in FY 1999. Onsite inquiries increased 20 percent to 56,737, in large part due to initial questions about the new online public access catalog and onsite circulation module of the Voyager Integrated Library System. The number of e-mail inquiries increased 47 percent to a total of 40,659.

In early FY 1999, LO completed the consolidation of first-line response to inquiries from remote requesters in the Reference and Customer Services Section. The CustomerQ software is now used to track both e-mail and telephone inquiries, allowing better analysis of questions received and feedback to the developers of NLM products and services. Sixty-five "Frequently Asked Questions" were added to the "Contact NLM" section of NLM's Web pages during the year. Simplification of the NLM telephone tree and the use of the same staff to answer both telephone and e-mail inquiries dramatically improved NLM's telephone service. By the end of the fiscal year, 97 percent of phone calls were answered without ever being in a queue. The LO and contract employees who handle first-line service were able to resolve 80 percent of the e-mail and telephone inquiries; 20 percent were referred to experts in other LO or NLM units.

#### **Outreach**

Many LO programs are designed to increase awareness and use of NLM's services by health sciences librarians and other information providers, health professionals, researchers, and the general public. LO coordinates the National Network of Libraries of Medicine (NN/LM) which works to equalize access to health information services and technology for health sciences libraries, health professionals, and the general public throughout the United States; develops major exhibitions and other programs in the history of medicine; and conducts a range of training programs for health sciences librarians. Many LO staff members give presentations and demonstrations at professional meetings and write articles to highlight the NLM programs and services.

#### *National Network of Libraries of Medicine*

The goal of the NN/LM is to provide U.S. health professionals, researchers, educators, administrators, and members of the public with timely, convenient access to biomedical and health information resources. The NN/LM strives to ensure equal access to up-to-date information irrespective of the user's location or institutional affiliation. The network includes more than 4,500 health sciences libraries, including hospital and academic medical center libraries, located throughout the country. LO's NN/LM Office oversees the network programs that are coordinated and administered by eight Regional Medical Libraries under contract to NLM. (See Appendix 1 for a list of the RMLs.) The RML contracts are re-competed every five years. LO began work on the request for proposals for the 2001-2005 contracts which will be issued in FY 2000. The new statement of work will add an emphasis on outreach to the general public and will provide increased flexibility to customize network programs to fit regional requirements.

Internet connectivity eases access to many high quality information sources, so NLM and the RMLs have focused on ways to improve connectivity for health sciences libraries and for health professionals. In FY 1999, NLM and the RMLs conducted a follow-up survey of NN/LM members that reported no Internet connectivity in FY 1997 or had joined the network since that time. The results indicate that 93 percent of all NN/LM members and 91 percent of hospital library members now have Internet access, up from 79 percent and 71 percent respectively in FY 1997. The RMLs will continue to assist in connecting

unconnected—and improving access for underconnected—network members.

The NN/LM is a core component of NLM's outreach program. The RMLs and other NN/LM members develop and conduct many special projects to reach underserved health care professionals in both rural and inner city areas. In FY 1999, the Library funded 76 special NN/LM projects to improve health professionals' access to information. Examples include:

- Michigan State University is partnering with the Michigan Public Health Institute to provide computer equipment and Internet training for the staff of 49 Michigan county public health departments and three rural hospitals.
- Idaho State University is providing Internet access and training for five community and migrant health center clinics in southeastern Idaho.
- Health professionals who provide health care for underserved Hispanic populations in 14 counties in Northwest Ohio are the targets of a project funded at Medical College of Ohio. Internet connections and training in accessing health information resources will be provided.
- Montana State University will assist tribal college librarians in providing information services to health care professionals in the Indian Health Service.
- Boston University Medical Center, in collaboration with the Boston Public Health Commission, will provide information skills training on HIV/AIDS resources to staff at Title I and City of Boston Prevention, Education and Care funded programs.
- The Southwest Georgia AHEC will provide hands-on training workshops in the use of PubMed and CDC Wonder for public health professionals on the staffs of 14 county public health offices.
- The University of Connecticut Health Center Library will upgrade the Internet connections for 22 district and municipal health departments and 14 part-time health departments in Connecticut.
- State and local public health professionals in the Dallas, Denton, and Fort Worth areas will receive intensive Internet training as part of the project at the Texas Woman's University Libraries. Onsite resource demonstrations will be conducted using a mobile computer lab.
- The Houston Academy of Medicine will work with the Houston Department of Health and Human Services, the Houston Public Library, and the Harris County Public Library to develop an outreach awareness campaign. The campaign will acquaint public health workers with the resources of NLM and CDC. A special public health homepage with links to NLM and CDC, and local, state and regional sites will

be developed.

A number of the projects targeted toward public health professionals resulted from Partners in Information Access for Public Health Professionals, a joint effort of NLM, the NN/LM, the Centers for Disease Control and Prevention, the Health Resources and Services Division, the Association of State and Territorial Health Officials, the National Association of City and County Health Officials, and the Public Health Foundation. The NN/LM Office works with the National Information Center for Health Services Research and Health Care Technology (NICHSR) and the Specialized Information Services Division to coordinate NLM's participation in this program, which is designed to improve access to advanced information technology and information services for practicing public health professionals. The partnership also led to a joint NLM/CDC satellite broadcast on Accessing HIV/AIDS Information Resources on February 11, 1999 and the addition of initial chapters of the Guide to Community Preventive Services to NLM's HSTAT (Health Services and Technology Assessment Text) database.

The NN/LM also plays a key role in NLM's new initiative to improve the general public's access to health information. An additional staff position has been added to LO's NN/LM Office and to each RML to assist in developing outreach programs to the general public. The NN/LM Office directed the FY 1999 pilot project to assist public libraries in providing health information services to their users. A joint effort of NLM, the Friends of the NLM, the Kellogg Foundation, the Medical Library Association, the Public Library Association (a division of the American Library Association), this project paired 40 public libraries or library systems in nine states and the District of Columbia with NN/LM health sciences libraries which could provide training and back-up reference and document delivery assistance. Pilot project participants became users and testers of NLM's new MEDLINEplus Web site, as well as other NLM services. The results of the pilot, which ran from October 1998 to June 1999, were assessed by a combination of reports, conference calls, site visits, focus groups, and NLM service use statistics. Although it proved difficult to isolate and measure the precise impact of the pilot project, participating public libraries were enthusiastic about the training received, were more confident in responding to health information questions from their patrons, and found the expanding MEDLINEplus Web site increasingly valuable in responding to these questions.

Experience gained from the pilot project was factored into a request for proposals issued to all NN/LM members for special outreach projects to improve the public's access to health information. The request encouraged health sciences libraries to collaborate with a range of players who can provide health information to the public, including health professionals, public health authorities, public libraries, schools, churches, and other community-based organizations. NLM received 75 proposals in response to this solicitation and will fund more than 40 projects in FY 2000.

The RMLs and other NN/LM members conduct the majority of exhibits and demonstrations of NLM products and services at health professional and other meetings around the country. LO staffs some of the exhibits in the Washington, D.C. area, at the annual meetings of the Medical Library Association, and at some distant meetings focused on health services research, public health, and the history of medicine. In FY 1999, NLM and NN/LM services were displayed at 201 exhibits at national, regional, and state association meetings across the U.S. The exhibit schedule is being expanded to include more meetings of public librarians and consumer groups. LO arranged for the development of new exhibit booths for use by all NLM program areas. The new modular booth, which debuted in a three segment island configuration at the Medical Library Association meeting in May, includes large plasma display screens for demonstrations and mini-training classes.

#### *Historical Exhibitions and Programs*

LO's History of Medicine Division periodically mounts major exhibitions in the NLM lobby and rotunda, with assistance from the Office of the Director, the Lister Hill Center, the Office of Administration, and the Office of Communications and Public Liaison. Designed for the interested public as well as the specialist, these exhibitions are part of NLM's outreach program. *Breath of Life*, the largest exhibition NLM has ever mounted, opened on March 23, 1999. HMD directed the development and production of this complex exhibition on the history of asthma and the current state of knowledge about the disease in collaboration with the National Heart, Lung, and Blood Institute, the National Institute of Allergy and Infectious Diseases, and the National Institute of Environmental Health Sciences. Dr. Sheldon Cohen served as senior advisor to the exhibition; the visiting curators were Robert Aronowitz, Carla Keirns, and Dot Sparer. The exhibition employs innovative interactive

displays developed by the Lister Hill Center, as well as historical objects and illustrations. The Friends of the NLM sponsored a special opening program and reception on the evening of March 22 for the NLM staff and invited guests, including residents and their families from the Children's Inn at NIH. Participants in the opening program included Jackie Joyner Kersee and Nancy Hogshead, Olympic gold medal winners and asthma sufferers, and the "Sesame Street" characters Rosita and Luis, who along with Dani, a special Muppet with asthma, performed part of a new bilingual multimedia childhood asthma educational campaign "A Is For Asthma." An online version of the exhibition is available on NLM's Web site.

HMD installs "mini-exhibits" each year in the exhibit cases at the entrance to the HMD reading room. *That Girl There is Doctor of Medicine*, commemorating the 150th anniversary of Elizabeth Blackwell's graduation from medical college, opened in January 1999. *Classics of Traditional Chinese Medicine* from the National Library of Medicine Collection opened in September 1999. HMD sponsors a series of seminars by historical scholars as well as special lectures to celebrate African American History Month and Woman's History Month. This year Dr. David McBride lectured on "The African-American Medical Experience; Perspectives and Prospects" on February 11, 1999, and Dr. Regina Morantz-Sanchez presented a talk on "Dr. Mary Dixon Jones on Trial: Historical Reflections on the Problem of the 'Difficult Woman'" on March 24, 1999. HMD staff members presented historical papers and lectures at professional meetings throughout the year and also published the results of their scholarship in books, chapters, articles, and reviews.

#### *Training Programs for Health Sciences Librarians*

LO develops online services training programs for health sciences librarians and other search intermediaries and oversees the activities of the NLM-funded National Online Training Center at the New York Academy of Medicine. LO also directs the NLM Associate Fellowship program for post-masters librarians and a continuing education program in health services research and related topics for practicing health sciences librarians.

In FY 1999, LO and the National Online Training Center taught 740 experienced searchers how to transition from ELHILL to PubMed and another 431 students the basics of searching PubMed. The training manuals for these courses are now available on the NN/LM Web site. NLM is moving toward an

interactive Web-based approach for some online training. Interactive courseware is now under development for the new version of PubMed.

The eight participants in the 1998/99 Associate Fellowship were the first who were eligible for the optional second year of the program, which will be spent in an academic medical center, hospital, or other health-related institution. The purpose of the second year is to give librarians experience in working with multi-disciplinary teams to integrate library and information services into the patient care, research, or education missions of a parent institution. Four of the 1998/99 Associates elected a second year program: two at the University of North Carolina and one each at the Medical College of South Carolina and the University of Louisville. The other four Associates accepted positions at the University of Alberta and the Alberta Heritage Foundation for Medical Research, the Naval Surface Warfare Center, under contract to the University of Arizona, and at NLM. Seven 1999/2000 Associate Fellows began the program in September. Four are members of minority groups.

HMD assisted NICHSR in identifying an historian to write a script for a half-hour video history of health services research to be used in NLM's introductory course on health services research for health sciences librarians. The video will draw heavily on the library of oral history transcripts and video interviews that LO has been developing as part of its efforts to document the history of health services research. Both the introduction to health services research course and a new health services research statistics class commissioned by NLM will be taught at the 2000 annual meeting of the Medical Library Association.

In celebration of National Medical Librarians Month in October 1998, LO and the Lister Hill Center created special Web pages highlighting the achievements of a number of health sciences librarians in the National Network of Libraries of Medicine.

### **Health Informatics Activities**

In addition to its basic service activities, LO staff members represent NLM on several initiatives designed to promote more effective health applications of advanced computing and communications technologies. In FY 1999, LO continued to serve on the Department of Health and Human Services (HHS) Health Data Standards Committee that is overseeing the implementation of the administrative simplification provisions of the Health Insurance Portability and Accountability Act of 1996. LO contributed substantially to the codes and classifications sections of the proposed rule on administrative transactions and assisted in analyzing and resolving comments received on the proposal. LO initiated and now directs a contract jointly funded by HHS-Department of Defense-Department of Veterans Affairs for the continued development and free distribution of LOINC (Logical Observations: Identifiers, Names, Codes), a detailed clinical coding system that will be recommended for use in the claims attachment standards.

NICHSR is the lead NLM component working with the Agency for Health Care Policy and Research to arrange a joint NLM/AHCPR workshop on advanced training for research at the intersection of medical informatics, health services research, and public health. The invited workshop, which is scheduled for January 2000, will involve representatives from the informatics and health services research training programs funded by NLM and AHCPR.

LO staff members serve as project officers on telemedicine evaluation contracts funded by NLM's Office of High Performance Computing and Communications. NICHSR also directs a research project designed to measure the impact of guidelines for meta-analyses on the quality of meta-analyses published in several prestigious journals.

**Table 1****Growth of Collections**

<i>Collection</i>	<i>Previous Total (9/30/98)</i>	<i>Added FY 1999</i>	<i>New Total (9/30/99)</i>
<i>Book Materials</i>			
<i>Monographs:</i>			
Before 1500 .....	578	0	578
1501-1600 .....	5,811	3	5814
1601-1700 .....	10,129	4	10,133
1701-1800 .....	24,472	5	24,477
1801-1870 .....	41,147	8	41,155
Americana .....	2,341	0	2,341
1870-Present .....	659,437	8,577	668,014
Theses (historical) .....	281,794	0	281,794
Pamphlets .....	172,021	0	172,021
Bound serial volumes .....	1,129,701	26,812	1,156,513
Volumes withdrawn .....	(66,481)	(2,254)	(68,735)
Total volumes .....	2,260,950	33,155	2,294,105
<i>Nonbook Materials</i>			
<i>Microforms:</i>			
Reels of microfilm .....	95,920	6,840	102,760
Number of microfiche .....	396,580	9,842	406,422
Total microforms .....	492,500	16,682	509,182
Audiovisuals .....	63,782	1,151	64,933
Computer software .....	1,408	112	1,520
Pictures .....	56,601	83	56,684
Manuscripts .....	2,729,072	129,885	2,858,957
Total nonbook .....	3,343,363	147,913	3,491,276
<b>Total book and nonbook .....</b>	<b>5,604,313</b>	<b>181,068</b>	<b>5,785,381</b>

**Table 2****Acquisition Statistics**

<i>Acquisitions</i>	<i>FY 1997</i>	<i>FY 1998</i>	<i>FY 1999</i>
Serial titles received .....	22,378	22,247	22,433
<i>Publications processed:</i>			
Serial pieces .....	150,713	146,921	123,823
Other .....	21,422	21,642	14,418
<b>Total .....</b>	<b>172,135</b>	<b>168,563</b>	<b>138,241</b>
<i>Obligations for:</i>			
Publications .....	\$5,282,348	\$5,266,996	\$5,370,797
(For rare books) .....	(\$267,221)	(\$251,293)	(\$292,603)

**Table 3****Cataloging Statistics**

	<i>FY 1997</i>	<i>FY 1998</i>	<i>FY 1999</i>
Completed Cataloging.....	20,296.....	18,803.....	14,396.....

**Table 4****Bibliographic Services**

	<i>FY 1997</i>	<i>FY 1998</i>	<i>FY 1999</i>
Services.....	FY 1997.....	FY 1998.....	FY 1999.....
Citations published in MEDLINE*.....	519,012.....	411,921.....	434,525.....
For <i>Index Medicus</i> *.....	499,794.....	388,022.....	421,423.....
Journals indexed for <i>Index Medicus</i> .....	3,211.....	3,302.....	3,394.....
Abstracts entered.....	398,576.....	312,064.....	338,435.....

\* *Figures for FY 1997 reflect reduction of a backlog carried over from FY 1996.*

**Table 5****Circulation Statistics**

<i>Activity</i>	<i>FY 1997</i>	<i>FY 1998</i>	<i>FY 1999</i>
Requests Received.....	630,190.....	694,281.....	751,732.....
Interlibrary Loan.....	353,408.....	374,791.....	396,516.....
Onsite.....	276,782.....	319,490.....	355,216.....
Requests Filled:463,011.....	523,081.....	570,966.....	
Interlibrary Loan.....	252,830.....	275,588.....	301,073.....
Photocopy.....	239,575.....	264,301.....	291,743.....
Original.....	11,670.....	10,167.....	8,229.....
Audiovisual.....	1,585.....	1,120.....	1,101.....
Onsite.....	210,181.....	247,493.....	269,893.....

**Table 6**

## Online Searches—All Databases

	<i>FY 1997</i>	<i>FY 1998</i>	<i>FY 1999</i>
Total online searches.....	21,000,000	104,000,000	191,000,000

## Table 7

### Reference and Customer Services

<i>Activity</i>	<i>FY 1997</i>	<i>FY 1998</i>	<i>FY 1999</i>
Offsite requests .....	21,297	27,070	54,542
Onsite requests .....	40,851	43,782	56,737
Total .....	62,148	70,852	111,279

## Table 8

### History of Medicine Activities

<i>Activity</i>	<i>FY 1997</i>	<i>FY 1998</i>	<i>FY 1999</i>
Acquisitions:			
Books .....	76	108	170
Modern manuscripts .....	750,000	274,530	129,885
Prints and photographs .....	306	849	1,773
Historical audiovisuals.....	36	94	114
Processing:			
Books cataloged.....	558	193	58
Modern manuscripts cataloged .....	0	0	0
Pictures cataloged .....	0	0	83
Citations indexed .....	2,430	1,516	1,022
Public Services:			
Reference questions answered .....	13,701	12,387	14,050
Onsite requests filled .....	6,313	3,733	3,672

# SPECIALIZED INFORMATION SERVICES

*Melvin Spann, Ph.D.*  
*Associate Director*

The Toxicology and Environmental Health Information Program (TEHIP), originally known as the Toxicology Information Program, was established more than 30 years ago within the National Library of Medicine in the Division of Specialized Information Services (SIS). Over the years TEHIP has evolved to provide for the increasing need for toxicological and environmental health information, taking advantage of new computer and communication technologies. Such mechanisms have enabled us to provide more rapid access to a wider audience. Our development of novel search capabilities means that users need not have extensive search knowledge and thus allows data to be relayed to them more efficiently. Finally, we have moved beyond the bounds of the physical National Library of Medicine, exploring ways to point and link users to relevant sources of toxicological and environmental health information wherever these sources may reside. This is being accomplished primarily through the TEHIP and AIDS Web sites developed and maintained by SIS. Continuous refinements and additions are made to allow easy access to a range of information collected by this Division. Significantly, our usage increased over the past year as we made access to all toxicology data free over the Internet. All access to NLM's AIDS/HIV information resources has been free for several years.

In FY 1999 SIS reexamined the scope and coverage of current programs, proposed new opportunities to enhance SIS information services, and investigated emerging areas, including the application of new multimedia technology in the delivery of information services. This examination has been guided by an Institute of Medicine (IOM) evaluation of the TEHIP Program in a report, *Toxicology and Environmental Health Information Resources: the Role of the National Library of Medicine*, released in the spring of 1997. Another IOM committee was convened in FY 1998 to prepare a follow-on report which has also been instrumental in our planning process. The report of this committee was published in 1999 as *Internet Access to the NLM's Toxicology and Environmental Health Databases*.

## Resource Building

The wide range of resources related to toxicology and environmental health information and AIDS/HIV information include many databases that are created or acquired as well as other services and projects.

The Hazardous Substances Data Bank (HSDB) continues to be a highly used resource, averaging over 6,000 searches each month. Increased emphasis was placed on providing more data on human toxicology and clinical medicine within HSDB, in keeping with earlier recommendations of the Board of Regents' Subcommittee on TEHIP. Changes to the composition of the Scientific Review Panel were made to accommodate the shift in content emphasis. Newer sources of relevant data are being examined for incorporation into new and existing data fields within the current 4,545 HSDB records. Because of increased staff efforts, more records are being processed through special enhancements, including source updates from various peer-reviewed files. These enhancements are being made possible by a customized Windows-based PC workstation with enhanced file-building features.

CHEMID (Chemical Identification File) is an NLM online chemical dictionary that contains over 344,000 records, primarily describing chemicals of biomedical and regulatory importance. It also contains an important set of regulatory data, collectively known as SUPERLIST. Over 13,000 records are augmented with the name and an indication of source for chemicals mentioned in one or more of 31 lists, e.g., the Department of Transportation Hazardous Materials List, and the Priority List of the Agency for Toxic Substances and Disease Registry (ATSDR). These data allow users to determine if a chemical is mentioned on a given list and under what name, as well as to search for chemical classes on these lists. During FY 1999, an extensive quality control effort was made on data in this file and new nomenclature data was added to enrich the file content.

TOXLINE (Toxicology Information Online) is an NLM online bibliographic retrieval service produced by merging "toxicology" subsets from some 18 secondary sources. TOXLINE and its backfile, TOXLINE65, contain data from sources that do not require royalty charges based on usage. At the end of FY 1999 TOXLINE and its backfile contained over two and one half million citations. Explorations of new ways to build TOXLINE are in process as NLM continues its move to distributed systems.

DIRLINE (Directory of Information

Resources Online) is NLM's online directory of resources including organizations, databases, bulletin boards, as well as projects and programs with special biomedical subject focus. These resources provide information to users which may not be available from one of the other NLM bibliographic or factual databases. DIRLINE continues to receive a high level of use through a new interface, which became public in October 1999. This new interface supports direct links to the web sites of the organizations listed in the database, as well as direct e-mail connections. Providing direct links for users facilitates ease of access for consumers as well as for health professionals. The quality of the database continues to improve as duplicates are eliminated through changes in policy and streamlining of maintenance. Health Hotlines, the popular publication of health-related toll-free telephone numbers, has a new web version that also indicates the availability of Spanish-speaking customer service representatives and Spanish language publications from the resources listed.

The Toxic Chemical Release Inventory (TRI) series of files now includes three online files, TRI95 through TRI97. These files remain an important resource for environmental release data and continue to attract new users. Mandated by the Emergency Planning and Community Right-to-Know Act (Title III of the Superfund Amendments and Reauthorization Act of 1986), these EPA-sponsored databases contain data on environmental release data to air, water, and soil for over 600 EPA-specified chemicals.

The Chemical Carcinogenesis Research Information System (CCRIS) continues to be built, maintained, and made publicly accessible by NLM. This data bank is supported by the National Cancer Institute (NCI) and has grown to over 8,000 records. The chemical-specific data covers the areas of carcinogenesis, mutagenesis, tumor promotion and tumor inhibition.

The Integrated Risk Information System (IRIS), EPA's official health risk assessment file, continues to experience high usage and be very popular with the user community. EPA has had a version of IRIS on the agency's Web page since 1996, and as we move to Web access we will consider how best to integrate our Web service with what EPA provides. IRIS now contains 535 chemicals.

The GENE-TOX file continues to be built and updated directly on TOXNET by EPA scientific staff. This file contains peer-reviewed genetic toxicology (mutagenicity) studies for about 3,200 chemicals. GENE-TOX receives a high level of interest among users in other countries.

The Registry of Toxic Effects of Chemical Substances (RTECS) is a data bank based upon a National Institute of Occupational Safety and Health (NIOSH) file by the same name which NLM has restructured and made available for online searching. With our move to free Internet access to all databases, NIOSH requested that we no longer include RTECS on our system.

The Developmental and Reproductive Toxicology (DART) database now contains over 43,000 citations from literature published since 1989 on agents that may cause birth defects. DART is a continuation of the Environmental Teratology Information Center backfile (ETICBACK) database, which contains almost 50,000 citations to literature published from 1950-1989. DART is funded by NLM, the Environmental Protection Agency, the National Institute of Environmental Health Sciences and the FDA's National Center for Toxicological Research and is managed by NLM.

The Environmental Mutagen Information Center (EMIC) database contains over 23,000 citations to literature on agents that have been tested for genotoxic activity. A backfile for EMIC (EMICBACK) contains over 75,000 citations to the literature published from 1950 to 1991. EMIC is funded by the Environmental Protection Agency and the National Institute of Environmental Health Sciences and managed by NLM.

### **Resource Access**

The SIS Web server provides a central point of access for the varied programs, activities, and services of the Division. Through this server (<http://sis.nlm.nih.gov>) users can access interactive retrieval services in toxicology and environmental health or AIDS/HIV information, find program descriptions and documentation, or be connected to outside related resources. During FY 1999, a redesigned SIS web site was made available to the public. The SIS web site now incorporates information about SIS in general, as well as toxicology and environmental health (<http://sis.nlm.nih.gov/tehip.htm>) and AIDS information (<http://sis.nlm.nih.gov/hiv.htm>). Both the toxicology and environmental health and AIDS web pages provide links to NLM outreach activities in these subjects, access to NLM databases, links to selected web sites in these subjects, as well as tutorials, fact sheets, and other publications produced by SIS. Over 7,000 users visit the SIS web site weekly and view approximately 40,000 pages.

### *Toxicology Data Network (TOXNET)*

The Toxicology Data Network (TOXNET), NLM's computer system providing access to the majority of its toxicology files, has moved from a networked microprocessor environment to a UNIX-based platform (Solaris Version 2.6) on a SUN Enterprise 3000 computer. Concurrent with this transition to improved hardware and software was the announcement that TOXNET access would be free via an easy-to-use Web interface. In FY 1999, SIS debuted a new search interface to access all of the SIS toxicology and environmental health databases. This new search interface (<http://toxnet.nlm.nih.gov>) allows users to easily search HSDB, TOXLINE, CCRIS, Gene-Tox, DART, EMIC, IRIS, and TRI. Based on recommendations from the Institute of Medicine, users are presented with a basic search screen with just a single input box for searching. Most relevant records are displayed first and all search terms are highlighted. The search system also assists users in searching for drugs or chemicals by automatically adding synonyms and CAS Registry Numbers. This feature enhances retrieval across databases, as well as simplifying search strategies for users. Users also have several options to sort records and to print and download all retrieved information. Custom search screens for more sophisticated users are under development.

### *Internet Grateful Med (IGM)*

Near the end of FY 1998 access to TOXLINE and ChemID was added to IGM, where access to DIRLINE, the AIDS/HIV databases, MEDLINE, and many other NLM databases was already being provided. This route of access (<http://igm.nlm.nih.gov>) is free to all users.

### *Chemical Structure Server*

The chemical structure server has evolved from a mechanism to provide structure searching for chemicals covered by SIS databases to a system for integrating chemical dictionary record building and structure searching. This system uses special molecular searching programs and includes a prototype database for construction of ChemID records. The chemical structure server was upgraded in 1999 to support ChemID*plus*, a new version of ChemID. It is available at <http://chem.sis.nlm.nih.gov/chemidplus/>. This new resource has all the nomenclature available in ChemID, plus over 57,000 chemical structures. Users may search by name or other identifier, as well as by chemical

structure, substructure, or similar structure. There are hyperlinked Locators in ChemID*plus* that automatically search other Web resources such as MEDLINE, HSDB, and TOXLINE and SUPERLIST resources for a given chemical. In 1999, making the file public, the beginning of a new online maintenance system was added to the system.

### **AIDS Information Services**

NLM has continued to refine its HIV/AIDS information services and make them more available to a wider audience. SIS staff led the development of the HIV/AIDS topic page on MEDLINE*plus*, identifying and organizing resources of specific interest to consumers. This page is a valuable addition to the redesigned NLM AIDS home page (<http://sis.nlm.nih.gov/hiv.htm>) which contains information about NLM's programs, access to the HIV/AIDS-related databases, and links to selected HIV/AIDS resources of a more technical nature.

NLM has continued its successful AIDS Community Outreach Program with 13 awards in FY 1999, bringing the total number of awards made under this program to 108. This year, for the first time, an award was made to another Federal agency, the AIDS Information Center, U.S. Department of Veterans Affairs, to place kiosks in the waiting rooms of infectious disease clinics. NLM-funded projects have ranged from the simple purchase of hardware and services to support a widely acclaimed web site (AEGIS), to the development of low literacy treatment fact sheets in English and Spanish, to supporting public libraries to establish AIDS resource centers within their institutions.

At the request of our partner PHS agencies, NLM assumed responsibility for project management of the AIDS Clinical Trials Information Service (ACTIS) and the HIV/AIDS Treatment Information Service (ATIS). NLM has been a partner in these projects since their inception, but now has increased responsibility. The ACTIS databases, AIDSTRIALS and AIDS DRUGS, are available through Internet Grateful Med, as well as on the Web.

NLM has provided training in the use of HIV/AIDS resources to a number of different audiences. As part of the collaborative Partners in Information Access for Public Health Professionals, SIS staff led the effort to conduct training via a two-hour satellite broadcast on February 11, 1999. This training, targeting those who work in local and state department of public health, included training in selecting high-quality, accurate resources, as well as a discussion of

the most valuable of those resources. NLM has entered into a relationship with the National AIDS Minority Information and Education Program (NAMIE) to provide training at regional and other meetings. In addition, NLM continues to provide training at a variety of Historically Black Colleges and Universities (HBCUs) to faculty, staff, and members of the local community.

### **Outreach / User Support**

SIS continues its support of the Toxicology Information Outreach Project. The objective of this initiative is to strengthen the capacity of HBCUs to train medical and other health professionals in the use of NLM's toxicological, environmental, occupational health and hazardous wastes information resources. In addition to providing workstations, training and free online access to HBCUs participating in the training development project, NLM has collaborated with the Agency for Toxic Substances and Disease Registry (ATSDR) to train representatives from additional schools in the use of NLM's valuable online resources. One of the training classes, hosted by Texas Southern University, included HBCUs and community-based organizations from the Lower Mississippi Delta. This class was jointly sponsored by ATSDR, NLM, and the Environmental Justice Office of the Environmental Protection Agency in support of the Mississippi Delta Project. Other classes with specific user group focus have been conducted in addition to our usual NLM-based training.

#### *User Support Computer-Based Activities*

In FY 1999, SIS developed TOXICOLOGY TUTOR III, third in a series of introductory level computer-based toxicology courses. It takes advantage of Windows' graphical user interface incorporating color graphics, photography, animation, and audio. The TOXICOLOGY TUTOR series is designed for the Internet, and is available on the SIS Web server. Additionally, SIS has developed computer-based demos for the Internet illustrating the TEHIP databases, ChemID, TOXLINE, RTECS, HSDB, CCRIS, IRIS,

TRI, TRIFACTS, GENETOX, EMIC, and DART. This and several other overviews are available on the SIS website.

#### *Alternatives to Animal Testing*

SIS continued to compile and publish references from the MEDLARS files that were identified as relevant to methods or procedures which could be used to reduce, refine, or replace animals in biomedical research and toxicological testing. Requests for these quarterly bibliographies have increased, as has the number of articles deemed relevant to the field. Bibliographies issued during the past four years are available on the Internet through the SIS Web server, and the primary distribution mechanism for this project is now the Internet.

#### *Other Specialized Services*

In addition to toxicologic data files, SIS is evaluating other areas for creating specialized factual and bibliographic databases. Resource allocations are being made to determine the feasibility of initiating more clinical medicine information products for public, health professional, and scientific audiences. One area that SIS is beginning to focus attention on is drug information. SIS is taking steps to begin a critical review of its role in organizing and disseminating drug information in various formats. Recognizing that drug literature is increasing exponentially, that the Food and Drug Administration is approving new drugs at a much greater rate with requirements for post approval research, and approximately 2,400 sole source drugs may still have research ongoing, the need for cataloging, disseminating information and identifying access points to relevant data sources has become especially important. Consumers are receiving drug information directly from pharmaceutical firms and numerous other sources, and one role that SIS is exploring is assessing the integrity and validity of such information. SIS is also proposing new initiatives that are timely and responsive to user needs, and will schedule focus groups to review such proposals to ensure that resources are utilized appropriately.

# LISTER HILL NATIONAL CENTER FOR BIOMEDICAL COMMUNICATIONS

*Alexa T. McCray, Ph.D.*  
*Director*

## **Introduction**

The Lister Hill National Center for Biomedical Communications was established by a joint resolution of the United States Congress in 1968 as a research and development division of the National Library of Medicine. Lister Hill Center research is carried out through several major programs, all sharing the purpose of improving health care information dissemination and use. We conduct our research by drawing on a diverse set of scientific fields and methods. Our researchers have backgrounds in medicine, computer science, library and information science, linguistics, engineering, and education. We are administratively organized into five major components, and our research activities are regularly reviewed by an outside advisory group, the Board of Scientific Counselors, whose members are drawn from the medical informatics community (Appendix 3).

The Audiovisual Program Development Branch conducts media development activities with three specific objectives. As its most significant effort, the branch supports the LHCBC's research, development, and demonstration projects with high quality video, audio, imaging, and graphics materials. From initial project concept through final project implementation, evaluation and reporting, including image creation, preservation, transfer and display, all forms and formats of imaging are supported. Consultation and materials development are also provided by the branch for the NLM's educational and information programs. With the mission requirement of the Library expanded to include effective outreach activities, the support that the branch provides to these programs continues to increase. From the use of optical media technologies to teleconferencing and World Wide Web design support, the graphics, video, and audio materials requirement has increased in quantity and diversified in format. The third area of concentration is the engineering of technical improvements applied to issues such as image quality and resolution, color fidelity, media transportability, media storage, and image communication.

In addition to technique development by the staff, the facilities and hardware systems must reflect state-of-the-art standards in a very rapidly changing field. High resolution video is a development area being explored that represents the future for improved electronic image quality. Multimedia systems and techniques, visualization and networked media are being pursued for the educational and the cost advantages that they offer. Three dimensional computer graphics, animation techniques, and photorealistic rendering methods have changed the tools and products of the graphics artists in the branch. Digital video and image compression techniques are central to projects being pursued to improve image storage and transmission. Included within the Branch is the office of the Public Health Service historian. Significant efforts of this office have been devoted to the PHS Bicentennial celebration, which was extended into 1999. The office also prepared biographical sketches and provided pictures of all fifteen former Surgeons General for addition to the web site of the Office of the Surgeon General. Information about Audiovisual Program Development Branch projects appears on the Branch Web server at <http://lhncbc.nlm.nih.gov/apdb/>.

The Cognitive Science Branch conducts research and development in computer and information technologies. Important research areas involve the investigation of a variety of techniques, including linguistic, statistical, and knowledge-based methods, for improving access to biomedical information. Branch members actively participate in the Unified Medical Language System project and collaborate with other NLM research staff in NLM's Indexing Initiative, whose goal is to develop automated and semi-automated techniques for indexing the biomedical literature. The Branch also conducts research in digital libraries and is currently collaborating with NLM's History of Medicine Division on Profiles in Science, a project to digitize collections of prominent biomedical scientists. During the past year, Branch staff have also worked on the development of a database of information on clinical trials.

The Communications Engineering Branch conducts applied research and development in image engineering and communications engineering, motivated by NLM's mission-critical tasks such as document delivery, archiving, automated data entry for the creation of MEDLINE records, as well as wide Internet access to biomedical multimedia databases, and future imaging applications in support of medical educational packages employing digitized radiographic, anatomic, and other imagery. In addition to applied research, the Branch also develops and maintains

operational systems for production. The Branch conducts research in the capture, storage, processing, online retrieval, transmission and display of both biomedical documents and medical imagery. Data types of interest include bitmapped bitonal document images, digitized color documents, digitized x-rays, color cryosection images from NLM's Visible Human collection, and motion video. Areas of active investigation center on document image analysis and understanding techniques, image compression, image enhancement, image feature identification and extraction, pseudo-grayscale rendition, image transmission and networks implemented via asynchronous transfer mode (ATM) and satellite technologies, optical character recognition (OCR) and man-machine interface design applied to automated data entry. Information on these projects appear in this report under the headings MARS, DocView, Biomedical Multimedia Database R&D, Advanced Medical Imaging Tools, Image Storage and Transmission, and Next Generation Internet. Information about Communications Engineering Branch projects appear on the Branch Web server at <http://archive.nlm.nih.gov/>.

The Computer Science Branch applies techniques of computer science and information science to problems in the representation, retrieval and manipulation of biomedical knowledge. Branch projects involve both basic and applied research in such areas as intelligent database systems, multimedia hypertext information delivery, and machine-assisted indexing for information classification and retrieval. Research issues include knowledge representation, knowledge base structure, knowledge acquisition, and the human-machine interface for complex systems. Important components of the research include embedded intelligence systems which combine local reasoning with access to large-scale online databanks. Other project work involves multimedia knowledge-based systems with interactive video capability and systems which search multiple databases with a single user query. Research staff include the teams which developed NLM's Internet Grateful Med and HSTAT (Health Services/Technology Assessment Text) programs and the team which annually produces the Unified Medical Language System Metathesaurus. Branch staff coordinate medical informatics training, including the eight-week NIH elective in medical informatics for third and fourth year medical students held each spring. Information about Computer Science Branch projects appears on the Branch Web server at <http://lhcnlm.nih.gov/lhncbc/organization/csb/csb.html>.

The Office of High Performance Computing

and Communications serves as the focal point for the NLM's High Performance Computing and Communications planning, research and development activities with federal, industrial, academic, and commercial organizations. The major activities of the office center around the Visible Human Project, the Telemedicine Program and the Next Generation Internet. During FY 1999 we have added an additional staff member with experience in anatomy, histology and the applications of computers to medical problems as well as Learning Center staff.

Lister Hill Center research activities involve both basic and applied informatics and fall into several broad research areas, and many of our research projects involve collaboration across our organizational units. Knowledge processing research includes language and information processing and knowledge based indexing systems. Information systems research includes consumer health informatics, database systems, digital library research, and medical education systems. Image processing research includes image segmentation, compression, and transmission methods and algorithms.

We provide ongoing informatics training to individuals at various stages in their careers. Trainees participate in collaborative research on our campus and have the opportunity to work closely with our staff in any of the research areas in which we are engaged. They have access to our on-site resources, facilities, and staff, and have the opportunity to pursue their research interests in the broad context of our research programs.

In support of our R&D activities, we host several laboratories, provide core system support, ensure system security, and conduct advanced network planning. In addition, we provide research support for Visible Human, telemedicine and health applications for the Next Generation Internet.

The most current information about Lister Hill Center activities can be found on our Web site at <http://lhncbc.nlm.nih.gov>.

## **Knowledge Processing**

### *Unified Medical Language System*

The Unified Medical Language System (UMLS) project conducts research to improve the retrieval and integration of biomedical information. The UMLS project continues to develop knowledge sources that can be used by a wide variety of application programs to overcome retrieval problems caused by differences in the usage of biomedical terminology across a wide range of information resources. The

Metathesaurus is the largest of the UMLS knowledge sources, representing multiple biomedical vocabularies organized as concepts in a common format. The SPECIALIST lexicon is a large syntactic lexicon of medical terminology, currently containing some 125,000 lexical items. The lexicon is accompanied by a set of lexical tools that are designed to help users abstract away from linguistic variations such as singular and plural forms of the same word, differences in punctuation and word order, and spelling variation.

The 2000 edition of the Metathesaurus will contain nearly 750,000 concepts. New editing and work management systems are facilitating better use of resources, better management and improved quality assurance. The porting of all Metathesaurus systems from the Ingres relational database management system to Oracle is on schedule, with production operations scheduled to switch over in the second quarter of 2000. The Metathesaurus 2000 release will be created in a new automated production system. The software and infrastructure changes should allow moving to a more frequent release schedule as active users exhibit a need for more current information. Examples include those using terminologies with frequent additions or changes, such as the MeSH Supplementary Concepts, clinical drug vocabularies, or other vocabularies with release schedules that do not synchronize with an annual cycle. The Metathesaurus development group has recently begun an interesting experiment with the government entities responsible for the Australian clinical modification of the ICD10 International Classification of Diseases. Editors in Australia have downloaded the NLM editing client and successfully used it on NLM Metathesaurus editing servers over the Internet. The NLM group looks forward to a synergistic collaboration for inclusion of the Australian clinical vocabulary and as a model for remote inter-group collaboration in vocabulary construction and maintenance.

The UMLS Knowledge Source Server is a tool for providing Internet access to the information stored in the UMLS Knowledge Sources. The purpose of the Knowledge Source Server is to make the UMLS data more accessible to users, including system developers. Access to the system is provided through a command-line interface, through an Application Programming Interface (API), and through the World Wide Web. The Web interface allows users to browse and explore the data and to see how those data are organized in the UMLS. Information about Metathesaurus concepts, semantic types and relations in the Semantic Network, and lexical items in the SPECIALIST lexicon can be found by querying the system. The command-line interface is best suited for

batch processing. Researchers can submit a list of terms to the server and retrieve a range of information about those terms, and they can further filter the results, limiting the result set, for example, by a particular attribute. The API allows developers at remote sites to embed calls in their application programs to the Knowledge Source Server, thereby accessing the UMLS data directly over the Internet.

#### *Medical Language Processing*

A text analysis system based on the SPECIALIST lexicon and the lexical tools is under development. The system is modular in design to allow for flexible use and continuous revision. The modules are servers that will be available to a variety of clients for a variety of uses. The system consists of the following modules: a tokenizer module to analyze text into tokens and label them; a sentence identification module to analyze text into sentences; a lexical look-up module to find lexical items in the text, a shapes module to identify items in the text that do not occur in the lexicon but have types recognizable from their form, and a parser module to assign phrase structure to the sentences of the text. The design of the parser module is strongly based on the syntactic information encoded in the SPECIALIST lexicon. After consideration of several grammar paradigms, Categorical Grammar was chosen as a grammatical system that expresses this lexically based approach to sentence structure.

#### *Semantic Knowledge Representation Project*

The Semantic Knowledge Representation project is concerned with reliable and effective management of the knowledge encoded in natural language texts. The project develops programs which provide usable semantic representation of biomedical text by building on resources currently available at NLM, especially the UMLS knowledge sources and the natural language processing tools provided by the SPECIALIST system. Two programs in particular, MetaMap and SemRep, are being evaluated, enhanced, and applied to a variety of problems in the management of biomedical information. MetaMap has been used in several recent research projects involving information retrieval in the biomedical literature. As part of the Indexing Initiative prototype under development, MetaMap Indexing uses MetaMap to contribute suggested indexing terms for MEDLINE citations. Combined with techniques for hierarchical indexing, MetaMap was applied to a subset of the HSTAT database in research exploring full-text retrieval in a

recent Library associate project. The algorithm underlying MetaMap provided the basis for the approximate matching methodology employed in the Large Scale Vocabulary Test. As an alternative to query expansion based on retrieval feedback when applying statistically-based systems to MEDLINE citations, MetaMap was used for associating Metathesaurus concepts with the original query. Experiments demonstrated that this methodology compares favorably with retrieval feedback.

MetaMap and SemRep together have served as the basis for several ongoing projects that explore the application of focused semantic interpretation to problems in biomedical information management. One such project explores the formative evaluation of a program for accurately identifying terminology associated with the coronary arteries as expressed in coronary catheterization reports. In another, two experiments were conducted which assess the ability of an enhanced version of SemRep to identify and characterize physico-spatial semantic relationships in these same reports. A separate application addresses the ability of MetaMap and a special version of SemRep to discover molecular binding relations in biomedical text.

#### *Indexing Initiative*

The indexing initiative project is investigating methods whereby automated indexing methods may partially or completely substitute for expert indexing of the biomedical literature by humans. The project is investigating concept-based indexing methods that go well beyond automatic word-based indexing. One method seeks to discover semantic relationships between phrases in text as a way of more accurately representing content. Another identifies Metathesaurus concepts in biomedical text and then maps these through a weighting algorithm based on linguistic and knowledge-based techniques to appropriate MeSH terms. One set of experiments has demonstrated the value of such concepts for automatic query expansion. Another investigation provides an enhanced representation of semantic content by ranking concepts assigned to MEDLINE abstracts on the basis of frequency of occurrence and specificity as measured by hierarchical depth in MeSH. The project was reviewed by the BoSC during FY 1999. Team members have developed a prototype indexing system that will undergo testing in FY 2000.

#### **Information Systems**

#### *Consumer Health Informatics*

Consumer health informatics is an important new research area for us. Increasingly people are turning to the Internet to look for answers to their health questions. This raises a number of research questions, including the type of content that should be created and how that content can be put into the appropriate medical context. As a result of recent legislation that required that the NIH create a database of clinical trials information, our research team has initiated the development of such a system. The database will be an extensive resource that provides patients, families, and members of the public with easy Web-based access to information about clinical trials, including information about which clinical trials are currently recruiting patients, where the trials are being conducted, what the design and purpose of the research study is, and what the criteria are for participating. An important feature of the database will be to provide access to other online health resources that help place clinical trials in the overall context of a patient's medical care.

#### *Internet Grateful Med and the NLM Gateway*

Released in April 1996 as a principal component of NLM's System Reinvention initiative, Internet Grateful Med (IGM) provides users with assisted searching in MEDLINE and 14 other NLM databases over the World Wide Web. IGM is "middleware," an intelligent gateway system designed to provide assisted searching to NLM's users with a consistent look and feel across multiple database systems as these retrieval systems evolve. For its approach to the problem of providing attractive Internet-accessible interfaces to legacy systems, IGM won one of several Federal Showcase Site awards at the Federal Webmaster Workshop in August 1996. By the end of FY 1999, usage of IGM was running at more than 5.4 million World Wide Web page views per month. IGM users from computers at 156,000 Internet addresses in dozens of countries made nearly 1.3 million searches in MEDLINE and other NLM databases during September 1999.

IGM searches MEDLINE on the PubMed retrieval system and searches 14 other databases on NLM's Elhill legacy system. The migration of the information in those databases to other NLM retrieval systems has led to the development of a new NLM Gateway which will replace Internet Grateful Med. Rather than requiring the user to choose a database and then searching in only one of 15 databases at a time as with IGM, the Gateway will search simultaneously in

all the databases to which it has access. Composite result sets are offered to the user, along with many new features for storing preferences and personalizing the interaction. A new full text retrieval system is part of the Gateway, offering a place to archive and access information which doesn't readily fit into other NLM retrieval systems.

The NLM Gateway has an object-oriented system architecture and is written in Java. It uses CORBA (Common Object Request Broker Architecture) and XML extensively. The initial version of the Gateway, to be deployed early in 2000, searches simultaneously in five databases or literature collections in four NLM retrieval systems. Once the Gateway is in place, additional functionality and access to other NLM retrieval systems will be added in phases.

#### *Smart Cards*

NLM sponsored or cosponsored several projects involving smart card technology during FY 1999. A smart card is a credit card sized plastic card with an embedded circuit chip. The chip can be a microprocessor with internal memory capable of running small programs, or simply a non-programmable memory chip. An NLM-sponsored project at the Concurrent Engineering Research Center (CERC) at West Virginia University uses smart cards for authentication and data storage in rural healthcare delivery. Both patient cards and health provider cards are used. NLM has been a cosponsor of the Western Governors' Association Health Passport Project. This project involves the storage of data from multiple Federal, state and local agencies on cards used by clients receiving benefits such as well child care, checkups, immunizations and food benefits.

Since June 1999 one of our Medical Informatics fellows has been exploring the possibility of storing on an advanced Cyberflex Java Card information specific to cardiac emergencies for patients known to be at high risk. If successful, this will provide a means of carrying particularly important information in a portable, updatable record for a specific patient population.

#### *Object Oriented Information Retrieval Project*

One hundred documents or chapters in the Health Services/Technology Assessment Text (HSTAT) collection have been linked to their associated abstracts on the National Guideline Clearinghouse web site. The links are bi-directional, with those abstracts also linking back to the HSTAT

documents. New HSTAT collections and data for FY 1999 include the Prevention Enhancement Protocols from the Center for Substance Abuse Prevention of the Substance Abuse and Mental Health Services Administration (SAMHSA/CSAP), a chapter from Community Preventive Guidelines by the Centers for Disease Control and Prevention, and a series of Evidence-Based Reports and Report Summaries from the Agency for Health Care Policy and Research (AHCPR, now the Agency for Healthcare Research and Quality). Four Treatment Information Protocols from SAMHSA/CSAT, two Living Document Guidelines in the HIV/AIDS Treatment Information Service (ATIS) collection, three AHCPR Technology Assessments, two Consensus Development Conference reports, and two Continuing Medical Education quizzes were also added to HSTAT during this fiscal year.

To make it easier for users to access the disparate information in the Health Services Research databases, OOIR project staff designed and implemented a subsidiary searchable system to provide a single point of entry to HealthSTAR, HSRproj, Dirline, and HSR Tools. The system, called HSRSearch, is available through NLM's National Information Center on Health Services Research (NICHSR) home page and as an option in HSTAT. It allows users to search all of the listed databases or any combination of them. The group implemented a batch search capability in HSTAT this year, designed for researchers and those with limited Internet access. The initial batch client allows users to email multiple queries to HSTAT. Unless flagged as ASAP queries, the emailed queries are processed when the overall system load is below a specific level. Results are returned by email to the user. A new build client, a web-based capability allowing contributors at remote sites to submit new documents to HSTAT or revise existing documents on HSTAT, is in its initial testing phase.

The OOIR group in the coming year will join the NLM Gateway development group, where it will focus on testing and experimenting with object-oriented database support, improved indexing methods for full-text in this environment, improved user navigation through large and complex documents, and appropriate search aids such as concept and vocabulary assistance.

#### *Digital Library Research*

Digital library research involves all aspects of creating and disseminating digital collections including proposed and adopted standards, emerging technologies and formats, copyright and legal issues, protection of

original materials, and permanent archival of digital surrogates. Research issues that are being addressed include long-term preservation of digital archives, innovative methods for creating and accessing digital library collections, the development of modular and open information environments, investigation of the role of well-structured metadata, and the exploration of different “points of view” on the same underlying data set.

In the fall of 1998 the Profiles in Science Web site was launched. This site is designed for scientists, scholars, and students, all of whom may gain an appreciation of the history of early scientific discoveries, and share in the excitement of the scientific enterprise. The collections have been donated to the NLM and contain published and unpublished materials, including books, journal volumes, pamphlets, diaries, letters, manuscripts, photographs, audio tapes and other audiovisual materials. The first collection featured on the site is a special collection of materials relating to the work of Oswald Avery (1877–1955), one of this country's first molecular biologists, whose findings proved that the genetic material is DNA. The second collection represents the papers of Joshua Lederberg (b. 1925), an American geneticist and microbiologist who received the Nobel prize in 1958 for his work in bacterial genetics. The most recent collection features the life and work of Martin Rodbell (1925–1998), an American biochemist and molecular endocrinologist who shared the Nobel Prize in 1994 for his discovery of G-proteins and the principles of signal transduction in cellular communication.

#### *MARS: Automating Data Entry into MEDLINE*

The goal of the Medical Article Record System, or MARS, is to develop automated systems for the extraction of citation and abstract data from medical journal articles to create bibliographic records for the MEDLINE database, based on research in document image analysis and understanding, database design, artificial intelligence using rule-based and artificial neural network approaches, graphical user interface design, speech recognition, and image processing. The production of MEDLINE records has traditionally been done by manual keyboarding. In response to a crisis in 1996 in this manual data entry, we developed a first generation system, MARS-1, that combined the keyboarding of citation data (journal name, date, author, title, affiliation, page numbers, etc.) with scanning and automatic text conversion by optical character recognition (OCR) of the abstracts which, if keyboarded, proves very labor-intensive. The first

generation system placed in production consists of workstations for scanning, reconciling (text verification) and keyboarding citations. In addition, the system has three servers: a network file server, a 5-engine OCR server and one to match double keyboarded data and link the citation to the abstract, all subsystems networked via a 100 Mbps LAN. Since February 1998, MARS-1 has been producing over 600 completed records/day, one-third of the NLM's total record production. The remaining two-thirds are produced by traditional keyboarding and also directly supplied in electronic, SGML-coded format by publishers.

Along with gradually improving and maintaining MARS-1 in routine production, staff conducted research toward more comprehensive automation, first focusing on the design of a database-centered and database-driven next generation MARS-2 architecture with the definition of over 130 tasks. The initial wave of tasks was concerned with the SQL Server database design, GUI issues and algorithms for automated processing. These included: the development of an entity relationship diagram, a data dictionary, and C++ database read/write classes; a daemon to interface the database to the OCR system; a scanner interface to the database and a new scanner GUI design incorporating a magnifying glass feature, a dialog box for confirming MRI (journal issue identifier), and a continue button for abstracts that span two pages; and comprehensive reconcile workstation software. Also, an initial top-level design for the recognition of Greek and other biomedical symbols is completed and work has begun on the design of a subsystem that will implement a feature extraction and matching classifier for these “special” symbols. A major part of this research toward more comprehensive automation also focuses on the development of algorithms and software modules for automated zoning, automatic field identification (labeling), and for reformatting field syntax.

#### *DocView*

The goal of the DocView Project is to conduct and apply R&D in document image processing to document delivery via the Internet. Key elements of this project besides its use of the Internet are that it addresses NLM's mission of providing document delivery to end users and libraries, and that it incorporates advanced digital imaging techniques. This research has resulted in a widely distributed client software, DocView 1.1, that enables end users to directly receive documents from Ariel systems or

websites. While Ariel, a product of the Research Libraries Group, is used by libraries and document suppliers routinely to send documents via Internet to similar organizations, there are few options for end users to directly receive them. The DocView project filled that niche with the release of the DocView client software in January 1998. By September 1998, this software was downloaded by 1,300 new users in 68 countries, and a year later (September 1999), there were over 5,000 registered users in 93 countries.

The DocView client software, which runs under any version of Microsoft Windows, enables an end user to receive documents over the Internet at the desktop, retain them in electronic form, view the images, organize the received documents into "folders" and "file cabinets", electronically bookmark selected pages, manipulate the images (zoom, pan, scroll), copy and paste images, and print them if desired. DocView also serves as a TIFF viewer for compressed images received through the Internet by other means, such as World Wide Web browsers. Users may receive document images either via Ariel FTP or Multipurpose Internet Mail Extensions (MIME) protocols. Using DocView, users may also forward documents to colleagues for collaborative work.

DocView has been publicized in a variety of ways: in addition to word of mouth by the original beta testers and other colleagues, people are finding out about DocView through AOL Netfind, Alta Vista, Yahoo and Excite which provide references to DocView's web pages. There is no indication of difficulties in either installation or usage as reported on the DocView Email List Server which enables subscribers to post help messages to other users. NLM is not providing any support for installation or use. The major users appear to be scientists receiving documents via Internet at their desktops from Ariel workstations at their libraries. This was the case, for example, for researchers at a malaria research site in Kilifi, Kenya who used DocView to receive medical articles from an Ariel system at NLM.

Four new interrelated DocView subprojects are being developed. The first is multimode delivery of multimedia information (HotMed). Based on the DocView software this system will provide documents on demand: it will handle scanned monochrome document images in addition to grayscale and color images, audio and video; it will use the Web as a delivery mode in addition to Ariel FTP and MIME email; and it will be scalable so that multiple HotMed-enabled machines can be used if the delivery load increases. The second subproject is Java-assisted document delivery. This is a Java applet based on

DocView software that can be delivered through the Web along with a document. The applet will provide document rendering and the other features of DocView, as well as platform- independence, unlike the current DocView which is available only for MS Windows-based computers. The third subproject is a 32-bit version of DocView providing new functionalities such as file conversion (e.g., from TIFF to PDF), image editing and handling of color and grayscale images, and a personal Web document library that a user can make available to colleagues and students. The fourth subproject, now in beta testing, is a document conversion server (DocMorph) which will enable users to upload image files for conversion to alternative formats, a recommendation by many DocView users who often find that the image files they receive cannot be used when they do not have suitable viewer software.

#### *Medical Education Activities*

The Learning Center for Interactive Technology (TLC) offers a wide range of computer-based information and educational technologies representing stand alone, intranet, Internet and World Wide Web applications in the health sciences. In FY 1999, over 900 individuals representing some fifty distinct groups visited TLC, bringing the grand total of visitors to over twelve thousand. In addition to the traditional constituency of health professional teachers, students, administrators, and researchers, TLC continued experiencing increasing numbers of visiting high school and college students. This increase has been due in large part to the great interest generated at the high school and college level by the Visible Human Project and the availability of a variety of Visible Human applications in TLC. Another contributing factor has been the inclusion of TLC in many of the general tours provided for visitors to the National Institutes of Health.

In addition to the enormous interest in multimedia applications of the Visible Human Project, visitors were also interested in: (1) Internet and intranet-based information and educational resources in medicine; (2) World Wide Web based instructional resources, particularly problem-based tutorials and simulations; (3) authoring tools for developing and evaluating network accessible instructional resources; (4) creating distributed learning environments in support of health professions education; and (5) using online collaboration tools.

Individuals and groups of visitors used the Learning Center to learn about and access a wide

variety of biomedical and educational resources available on the Internet and World Wide Web. Staff provided demonstrations and facilitated "hands-on" practice with new and emerging online technologies and resources and assisted visitors in reviewing existing Internet accessible multimedia resources in medicine, using the Center's Web server as a focal point. Workshops were conducted on how Internet, intranet, and related communication technologies can be used for effective learning. Tutorials on these topics were conducted in collaboration with the Radiological Society of North America. TLC resources were used by summer students and teachers from NLM's adopted school, Coolidge High, to develop materials and prepare their project presentations.

The Distributed Learning Environments project investigates innovative means for assisting health science institutions in their use of online distance learning technologies. An important goal of the project is to develop, test, and utilize collaborative strategies that institutions can use to successfully integrate a wide range of technologies. TLC staff and staff at the University of Missouri Health Science Center completed literature reviews of research in areas important to project implementation: adult learning; cooperative learning, problem-based learning, computer-supported collaborative work, and diffusion and adoption of innovation. Each review summarizes findings and contains a list of issues and questions for further research. The reviews were compiled into a distributed learning resource that was made available on the TLC web site. Staff developed a working prototype database to Internet accessible health professions education materials. The database, called EtherMed, is intended to be used as a research tool and a mechanism for sharing online courseware among health professions schools. Staff continue to experiment with the Lucent video conferencing technology in the TLC and are collaborating with NASA and the University of Alabama, Birmingham (UAB) to evaluate other conferencing tools and distance learning technologies. They are also working with UAB to do research involving the EtherMed database and video-conferencing tools.

FY 1999 was the beginning of an intramural research effort in telepathology. The research is designed to explore the level and nature of the interactivity and communications resources needed to carry on a successful clinical telepathology program. Earlier work on the Digital Microscopy System project has involved the indexing of a library of medical photomicrographs so that such materials can be available to medical practitioners, researchers and

students via the World Wide Web. This project will soon be integrated with the new research efforts in telepathology.

Lister Hill Center staff coordinated the production and implementation of each of the interactive components of the "Breath of Life" asthma exhibit, which opened March 22, 1999. Technologies utilized included DVD, CD-ROM, audio CD, laserdisc and videotape. A prototype interactive DVD program was developed featuring the exhibit. Utilizing existing digital media from the exhibit, a virtual exhibit experience is offered. A three-dimensional virtual space representing the exhibit structure within the NLM Lobby and Rotunda was also developed. Guided animation within this environment is featured in the DVD, as an example of potential use for a virtual environment. The prototype also reflects the potential for the delivery of high quality motion video embedded within an interactive program, and its effectiveness in representing large-scale exhibits. This project will be presented as a prototype for consideration as the NLM determines an efficient mechanism to travel the current and future HMD exhibits.

Several videotapes of NLM supported telemedicine projects were completed this past year. In May, sequences were videotaped at the University of Iowa, Iowa City and Ottumwa, Iowa. In June, the "BabyCare Link" project at Beth Israel Hospital, Boston, was taped. In July, at three Indian reservations in Washington, sequences were videotaped for a University of Washington "Tribal Connections" segment. During that trip, a segment for the Oregon Health Sciences University "Teledermatology" was videotaped in Hermiston, Oregon. The draft script for the Visitor Center Multimedia Program was received in August and is being reviewed. In September, videotaping was completed of the West Virginia University projects in the Huntington vicinity and elements of this were shown at a Capitol Hill Reception in September.

The Movement Disorders Video Database Project is a collaborative project with Yale University School of Medicine's Movement Disorders and Neurodegenerative Diseases Clinic, the Center for Advanced Instructional Media and the Biomedical Communications Department. This pilot effort establishes a digital video database of high quality, full-motion video of medical significance. Neurologically based movement disorders have been selected as subject matter which may be best characterized by video and sound. A production team traveled to New Haven, CT to videotape patients undergoing treatment at the Yale Movement Disorders

Clinic. Each patient was videotaped undergoing a specific set of diagnostic routines which are used clinically to diagnose and measure degrees of disease and progression. The data set currently contains 40 patients who have been diagnosed with either Parkinson's Disease, Huntington's Disease or Cervical Dystonia or Torticollis. In addition, 4 individuals without movement disorders were videotaped. Video segments of these individuals will enhance the full data set by providing "normals" for comparison. Yale principals Drs. Carl Jaffe and Ken Marek will utilize the data in their network-based Informatics Approach to Neurology of Movement Disorders.

An 8-minute video entitled "Frankenstein: Discovering the Secrets of Nature" was created for presentation at the Fifty Year Reunion of the Markle Scholars in Academic Medicine, in Phoenix, Az. In February 1999, the videotape was presented with the Communicator Awards "Crystal Award" (scoring 90+ points out of a possible 100). In March, it won a 20th Anniversary Classic Telly Award "Gold Medal Winner" that recognizes the best video productions of the last 20 years. These awards recognize individuals and companies in the non-network TV field whose talents and creativity achieve a high standard of excellence that elevates them above the best in the field and serves as a standard for the industry.

## **Image Processing**

### *Visible Human Project*

The Visible Human Project data sets are designed to serve as a common reference point for the study of human anatomy, as a set of common public domain data for testing medical imaging algorithms, and as a test bed and model for the construction of image libraries that can be accessed through networks. The Visible Human data sets are being made available through a free license agreement with the NLM. They are being distributed to licensees on DAT tape, and over the Internet at no cost. The data sets are being applied to a wide range of educational, diagnostic, treatment planning, virtual reality, artistic, mathematical and industrial uses by over 1300 licensees in 41 countries. They are used as a normal reference and as an aid in the diagnostic process. Programs under development will be used to educate patients about the need for and purpose of surgery and other medical procedures as well as to permit physicians to plan surgery and radiation therapy. The images from the Visible Human data sets are used in several prototype virtual reality surgical simulators. Educational materials

that make use of the Visible Human data sets are beginning to be used by students from kindergarten to practicing health care professionals. The data sets are being used to form the basis of interactive games to entertain as well as to educate. Automobile manufacturers now include passenger injury models based on Visible Human data to their vehicle crash simulation models. Engineers and physicists are creating models to quantify human exposures to various forms of electromagnetic radiation. The data provided by the Visible Human project are being used by mathematicians as an application for what were previously only theoretical mapping theories. Several artists are using the data set as the basis for new multi-media art forms.

During the past year a prototype anatomical image management system was developed using an object-oriented persistent storage engine. This prototype consists of two database systems; a data management system which performs image processing functions on input data; and an online database which provides Internet delivery of anatomical images. The database system provides two modes of web access, its anatomical browser supports the online display of low resolution images, and queries to its user database

return high resolution image records. Four types of image records are available; anatomical cross sections with labels, volume of interest, segmented masks, and rendered images. Presently, the user database stores 2D pixel images, and 3D Vogel structures for the thorax of the Visible Human male. These anatomical structures of the thorax were extracted from the Visible Human male 70mm color film data set. In addition, several support modules were developed to support the construction and display of visual products. A visualization module was implemented to produce surface rendered images using segmented masks which are also available for user development. The visualization module is a component of the data management system. A visual display module that processes and displays retrieved 2D and 3D anatomical images on user workstation was also developed.

During this past year, the quality assessment of a National Geographic duplicate of the 70mm film from the Visible Human Male was completed. Several slices from each block were scanned using the KODAK afs5057 scanner and were then compared to the scans of the original film. It had been determined that reviewing the digital scans provided the best way to assess the quality of the duplicate film. The contact duplicate of the 70mm film was determined to be of acceptable quality. In collaboration with Stokes Imaging, we have examined ways to digitize, duplicate and archive the 70mm film images of the Visible Human Female. The first step will determine the highest resolution digital image that can be acquired from the film using a customized imaging system built by Stokes Imaging. The next step is to convert the digitized images back to 70mm film and compare the quality of this method of duplication to the quality of duplicates produced by contact printing of the original 70 mm film. One hundred images will be sampled from the total of 5100 images in this first phase before proceeding to digitize and duplicate the entire set. Phase One of the digital duplication of the Visible Human Female 70mm film images was begun by producing test images from selected 70mm frames. Eight CD's containing these images were received for review in September and will be evaluated to determine the highest quality digital image that can be acquired from the film.

In collaboration with Yale University's Center for Advanced Instructional Media, a web-based teaching tool for head anatomy, with an emphasis on cranial nerves, is being developed. Volumetric data from the CT scans of the Visible Human Female were used to reconstruct 3D views of the skull. In addition, orthogonal views showing each cranial nerve were

reconstructed from volumetric RGB data. The anatomic structures in these 3D images were then labeled and added to the website.

Work continued on the graphical interface for the web browser used to access the database of the thorax anatomy. This graphical interface for a web browser enables users to interactively view rendered images of the thorax from four main areas: the thoracic cavity, the thoracic wall, the blood vessels, and the spine. Four composite images, which represent the main areas, were rendered and hot spots were added linking them to images of specific anatomic areas or structures. For example, a composite image of the thoracic cavity has links to images of the heart and lungs. These images were then further linked to more detailed images. The image of the lungs, for example, is linked to an image of the bronchial tree. A module that allows the user to find a specific area or structure using the anatomic terms of the thorax was developed and will utilize the rendered images that were created using a surface-rendering program. This module will consist of about 30 rendered images representing thoracic structures already labeled and an additional 50 images representing other thoracic structures that will be labeled in the future.

#### *Image Storage and Transmission Optimization*

This project involves research into compression and transmission techniques to improve access to, and delivery of, data-intensive biomedical images, with specific focus on the Visible Human (VH) color image set. The CCD captured male and female data in VH amounts to 55 GB, but the 70mm photographs taken during cryosectioning of the cadavers, currently being scanned at much higher resolution, will yield a total of about 235 GB. Since datasets of this size will strain both storage and transmission resources, this project is undertaken as an investigation of both compression and advanced communications techniques to alleviate these problems. Image compression has been the main focus of this project in 1999, and we have made progress in creating mechanisms for the access, retrieval and display of compressed images. The compression research has investigated both lossy and lossless techniques, since both classes have applicability: lossless techniques would serve the objective of retaining all the original image data in storage; however, since much higher compression ratios are possible with lossy techniques, these would serve our objectives of rapid transmission over the Internet. Studies conducted in lossy compression techniques included wavelet

transformation for image decomposition and both scalar and vector quantization for data reduction. For lossless storage, two sequential steps are being developed: automated segmentation of the anatomic region and fiducial marks from the background, and lossless compression by an adaptive arithmetic coding technique.

Progress was also made toward developing the tools for accessing, retrieving and viewing the compressed VH images. First, a Java applet, Image Retriever, was developed for selecting one or a range of slices clicking on a simplified body map created from the VH set, establishing the TCP/IP connection between client and server, and saving the compressed image in the user computer memory. Second, a Java application, Image Viewer, was created to allow the user to choose the image(s) to decompress, run the C code, and display the image in its raw form (the form it exists in the VH set) without requiring it to be in GIF or JPEG form.

#### *Biomedical Multimedia Database Research*

The goal of this program is to address fundamental questions that arise in the handling, organization, storage, access and transmission of very large electronic files in general and digitized x-rays in particular. A special focus is research into these topics as applied to heterogeneous multimedia databases consisting of both images and text. In collaboration with two other agencies, the National Center for Health Statistics (NCHS) and the National Institute of Arthritis, Musculoskeletal and Skin Diseases (NIAMS), Lister Hill Center staff assumed the role of technical manager and developer of prototype systems that served as testbeds to investigate: (a) image compression techniques, especially high yield lossy methods, and tools to interactively select compression parameters; (b) techniques to organize images and associated textual data for ready retrieval and use; (c) procedures and algorithms to implement transparent hierarchical storage using heterogeneous storage systems and media to match usage patterns; (d) tools to display x-ray images, read and record findings; and (e) multsocket transmission methods to segment large images and to send the pieces concurrently over multiple socket pairs to overcome the inefficiencies of conventional transmission protocols.

These research activities have relied on the availability of 17,000 cervical and lumbar spine x-rays collected during the second National Health and Nutrition Examination Survey (NHANES II), one of a series of nationwide surveys conducted by NCHS designed to provide a snapshot of the nation's health.

All 17,000 films were digitized with a Lumisys laser scanner resulting in images of approximate size 5 MB (cervical) and 10 MB (lumbar). These were quality checked and archived in our 144-platter optical jukebox, and provide the principal image test set for research.

The Web-based Medical Information Retrieval System (WebMIRS) is a research and development project whose goal is to exploit and evaluate Web technology for providing wide-area network access to multimedia biomedical databases, specifically those containing both text and digitized medical x-ray images. In addition to acting as an access mechanism, WebMIRS is also intended as a tool for data analysis by enabling query results to be imported into appropriate statistical packages. Specific objectives for WebMIRS include: (1) providing database access through any standard Java-enabled World Wide Web browser, such as Netscape or Internet Explorer, on any hardware platform supporting these browsers whether Unix computers, PCs, or the Macintosh; (2) relational database query capability through innovative Graphical User Interface design; (3) retrieval of x-ray images as well as text, when associated images exist; (4) capability to save results for review; (5) capability to export results to standard statistical tools (SAS, SUDAAN) for analysis; and (6) support for multiple databases.

To efficiently exploit the rich visual information in large biomedical multimedia databanks, we are developing a query by image content capability, critical for efficient information extraction from biomedical image data in the future. A necessary step toward such a capability is developing automated or semi-automated methods for extracting biomedical information from the archive of NHANES II digitized x-rays, for example, to segment and identify particular vertebrae in the spinal images. To this end, we are developing techniques to obtain basic geometric orientation relative to the anatomy contained in these images, for the purpose of eventual identification and measurement of features of interest to the biomedical research community

In joint research with radiologists and engineers at Georgetown University, we have acquired quantitative vertebral segmentation data for 550 images of the cervical and lumbar spines from the NHANES II survey. We first selected the images by searching the WebMIRS databases for subjects who met particular criteria for age and back pain symptoms, doing a batch search for the required image ID numbers, identifying the jukebox platters on which the images reside, and reading the images to hard disk after which they were

recorded on 8mm tape. A Georgetown University radiologist specialized in musculoskeletal x-rays displayed and recorded segmentation information from these images. To our knowledge, this is the first set of quantitative anatomical information obtained for a subset of the images from this survey. The data consists of (x,y)- coordinates for the six standard boundary points used in vertebral morphometry for every vertebra in these images. In addition, quantitative data was collected which describes the location of osteophytes, fused vertebrae, and, for each image, bounds of a rectangular subimage which approximates the smallest rectangular area containing each vertebra. This data is valuable for many reasons: (1) it provides the fundamental quantitative content for evaluating the usefulness of a database of spine anatomy measurements; from the basic data obtained, measures of vertebral sizes, intervertebral spacing, ratios of anterior/posterior vertebral heights, and degree of presence of osteophytes can be derived; all of these measures are of potential interest to spine researchers; (2) it provides the basis on which a priori spine geometry may potentially be inferred; a priori estimates of the location of one vertebra relative to another, for example, may prove useful in computer algorithms intended to estimate these locations with precision; and (3) it provides benchmark test data for evaluation of computer algorithms to locate spine anatomy.

#### *Advanced Medical Imaging Tools*

In concert with the research conducted in multimedia biomedical databases, and the increasing trend for medical information databanks to incorporate images, tools are being developed to easily search and retrieve such data over the Internet, to extract features from images, to enable content experts to display and read images, and to evaluate the returned image data against "gold standard" or reference images. The objective of this project is to address fundamental questions in the design of such tools required in the handling, organization, storage, access to, and transmission of very large electronic files in general and digitized x-rays in particular. Tools developed as part of this project are: a platform-independent digital online radiological atlas of the cervical and lumbar spine, a workstation to determine the scan resolution needed to effectively digitize hand x-rays, a quality control workstation to determine the quality of scanned hand x-rays from the NHANES III survey, and alternative packaging techniques for multimedia databases using consumer-level technology.

The growth in digital medical imaging

requiring expert interpretation has given rise to the need for convenient on-line digital reference tools, to assist in producing interpretations which conform to recognized standards. Inhouse research is contributing to this growing body of digital resources by building a digital atlas to allow for the standardized interpretation of cervical and lumbar spine images for the presence of osteoarthritis. Development of our atlas was originally motivated by the need to provide a reference tool for the standardized interpretation of a specific collection (NHANES II) of spine images acquired by NCHS. However, we have designed the atlas to be usable as a general reference tool on any Java-enabled platform including Sun workstations and IBM-compatible personal computers, with the capability of image display on a wide range of monitors. The atlas is being developed as both a downloadable Java applet and as a Java application.

We developed two high-resolution display workstations to support a digitization study prior to the scanning of the 10,000 NHANES III x-rays, currently being done. A key issue here was the minimum scan density (spatial resolution) required to read this class of x-rays when digitized. Each workstation consists of a Sun Sparc-20 machine and monitor, in addition to a high-resolution E-systems Megascan monitor. The workstations run on Java software, the dialog framework built entirely in Java's Abstract Windowing Toolkit, and a C++ image display module. The user sequentially displays a hand x-ray on the Megascan monitor, and, for this image, records a reading of the image for bone erosion on a form which is simultaneously displayed on the Sun monitor. The software records all of the data collected in a database table, and a report of the contents of this table can be requested by the user for print or display. To enable our NCHS colleagues to conduct quality control for the hand x-rays now being digitized, we developed quality control software and transferred it with associated equipment (Megascan monitor, external hard disk and a magneto-optical drive) to that agency, in line with an interagency agreement.

In order to evaluate the feasibility of inexpensive consumer-level software technology for the dissemination and portability of biomedical multimedia information, a prototype database is being created, consisting of text items collected in the NHANES II health survey, along with a subset of digital x-ray images collected in that survey. The text items in this database are chosen from the data collected in the demographic, anthropometric, and adult health questionnaire areas, and pertain to questions of pain in the cervical or lumbar spines. The images in the

database are a set of 550 images of the cervical and lumbar spines, half representing survey participants with self-reported back pain, and half without such pain. The images are available in two forms: (1) images with vertical and horizontal spatial resolution reduced by a factor of four; and (2) images at full spatial resolution, but limited to the spine vertebrae, the principal region of interest. In both cases, the images are of 8-bit grayscale depth, reduced from the original 12-bit depth. The feasibility of distributing this database on CD-ROM or DVD, once it is completed, is currently being evaluated. We expect this database to be completed in calendar year 2000.

### **Medical Informatics Training**

The Medical Informatics Training Program was established in 1996 with the goal of supporting medical informatics training opportunities at the LHCBC. The program provides participants an opportunity to train on site at the NLM and collaborate actively in LHCBC research in a variety of areas of medical informatics. Participants interact with outstanding professionals, have access to on-site resources and facilities, become familiar with and make significant contributions to research, gain an understanding of the Center's programs, and have the opportunity to make professional contributions to their fields. Potential areas of research for participants include digital library research, automated indexing techniques, vocabulary and thesaurus research, medical language processing, image processing, document analysis, belief networks, wide area network technologies, client/server design, database design, and computer-based learning.

During FY 1999, the MITP program had 25 participants from 10 states and seven foreign countries. Projects done by this year's students included World Wide Web-based systems, applications using the Unified Medical Language System, information retrieval projects, natural language processing, telemedicine, image processing, and image databases. Students worked closely with NLM staff and presented the results of their work to the NLM community.

The NIH Clinical Elective in Medical Informatics, one of 29 different rotations at the NIH, was again held at NLM in March and April 1999. The Center trained six third and fourth year medical students from U.S. and Canadian medical schools. The elective provided an overview of medical informatics through outstanding lectures both from invited speakers and from NLM staff and through two field trips to area medical facilities. Students had the opportunity to work

closely with NLM and NIH preceptors on independent research projects, presenting their work in a series of seminars at the end of the elective.

This year LHCBC continued to support minority students by participating in the college student summer employment programs of the Hispanic Association for Colleges and Universities (HACU) and the National Association for Equal Opportunities in Higher Education (NAFEO). The Center sponsored a HACU student from the University of Texas-Pan American in Edinburg, Texas and a NAFEO student from Fayetteville State University in Fayetteville, North Carolina. Both students had rewarding work experiences at NLM and an enriching experience living near the nation's capital.

### **Resource Support and Development**

#### *System Security and Advanced Network Planning*

Work during this year covered three areas: computer security, the Next Generation Internet, and internal networks and systems. Computer security concentrated on the continued deployment of authentication methods and the deployment of intrusion detection and system scanning processes. The implementation of the NLM connection to the Next Generation Internet was completed and our network was upgraded to match the higher speed capability of the NGI.

The deployment of intrusion detection and network scanning systems has been enlightening. The network scanning systems can scan both Windows NT and Unix systems and detect security problems. They have detected numerous Windows NT security problems which when rectified will help prevent security intrusion incidents. Within the Center, an authentication policy based on the SSH access protocol is being deployed. The SSH system encrypts all transactions and provides a secure authentication method independent of the Unix password system.

Work on our network has continued the expansion of the ATM (Asynchronous Transfer Mode) infrastructure. New Cisco 5500 switches have provided both ATM Lane processing capabilities and a comprehensive Fast Ethernet capability. The ATM Lane processing is now done on one of the 5500 switches, reducing the load on our routers. The Cisco 5500 switches provide both Layer 2 and Layer 3 switching capabilities to staff workstations. The Layer 3 switching increases overall throughput by doing inter-network routing on the switch rather than on the router.

### *Emerging Network Retrieval Protocols*

The project is exploring methods in support of the persistent naming of objects containing digital information. This work has included feature analysis and timing studies of systems based on the use of the Domain Name System (DNS), the Lightweight Directory Access Protocol (LDAP), the Persistent URL (PURL) methodology of the Online Computer Library Center (OCLC), the Handle system from the Center for National Research Initiatives (CNRI), Apache web server redirection, and ad hoc databases.

The group has also been exploring the problem of delivering program content in the character sets of different languages. One means of addressing this problem is the Unicode standard, which provides "a unique number for every character, no matter what the platform, no matter what the program, no matter what the language." A series of working Java-based demonstrations of Unicode use under the Unix operating system will help present the Unicode approach to development staff.

Project staff have investigated emerging mechanisms for the delivery of dynamic content over the World Wide Web, including methods referred to as "dynamic HTML." This has involved extensive use of the PHP server-side scripting language, the open source Apache web server, and the MySQL relational database system, with added elements of Java and JavaScript. Ten such systems have been developed and are in operation for our web services. One technical report and associated open source package has been released; nine more packages are in preparation for release. The released package demonstrates integration of ISDN telephone services with a World Wide Web client.

### *Next Generation Internet*

In March 1999 NLM was linked with an OC-3 (155 Mbps) connection to the high speed vBNS network, following a series of technical and administrative activities by Lister Hill Center staff with both MCI-Worldcom and Bell Atlantic. This connects NLM with over 88 major educational and research institutions on vBNS, in addition to peering with other high speed networks and therefore with a number of other national and international centers. Following the proposal evaluation to select a vendor for the OC-3 service to the vBNS node and arranging for NIH to be nominated as a vBNS Partner Institution, Lister Hill Center staff participated in the negotiations with MCI-Worldcom to resolve outstanding technical and

contractual issues. Concurrently, we investigated the value of moving from vBNS to vBNS+, and proceeded with the shift to vBNS+ having discovered that with the latter we would not be bound by the authorized use policy limiting our use of vBNS, and moreover would not entail changes in cost or term.

To implement a significant Next Generation Internet (NGI) application involving the interactive segmentation and indexing of the Visible Human images, planning began for the Trans-Pacific Digital Libraries Experiment, a collaboration between NLM and the Sapporo Medical University in Japan. A beta version of the Visible Human Viewer application developed by collaborators at Sapporo Medical University was obtained, and tested at NLM as well as with colleagues at NASA Goddard and NASA Ames Research Center. These sites will be instrumental in providing end-to-end connectivity for NLM and Sapporo, and the current test activities are designed to ensure reliable connectivity at different locations along the path. In connection with the Trans-Pacific Digital Libraries Experiment, Lister Hill Center staff participated in the NASA sponsored invitation-only "Bridging the Gap" workshop at NASA Ames Research Center in August 1999.

To complete the link between NLM and Sapporo, an arrangement has been negotiated with Teleglobe to provide last-mile' connectivity between the British Columbia Gigapop at Vancouver and the Lake Cowichan earthstation via DS3 service, with the understanding that NASA will procure this service. The expected North American path is now NLM-ATDnet-NASAGoddard-NREN-STARTAP-CA\*Net2-BC Gigapop-Lake Cowichan. Efforts are under way to obtain Intelsat certification for the Japanese earthstation to complete the link.

In order to conduct the technical evaluation part of the Trans-Pacific Digital Libraries Experiment, equipment was procured for our participation in a metrics program called Surveyor that is sponsored by the Internet Engineering Task Force. Surveyor involves linking a local PC to vBNS with software that allows the measurement of one-way delay and loss to other sites at several worldwide locations. The one-way delay measurement gives a more comprehensive view of actual link performance than the conventional roundtrip measurements typically done, and is important to identify particular factors barring achieving rated speeds over a link.

The Lister Hill Center served as a federal representative to the Maryland Governor's Task Force on High Speed Networks, to develop a comprehensive plan for bringing the state's network infrastructure in

line with the needs of the 21st century. This plan, completed and presented to the legislature, contains recommendations to: a) combine existing state resources to maximize the state's return on investment; b) use existing state owned fiber where available; c) use current right-of-ways the state possesses to add additional fiber in underserved regions such as the Eastern Shore, Western and Southern Maryland; d) provide equity of access to all regions of the state, and support multiple segments of our society; e) promote collaboration among businesses, educational institutions, governmental bodies and research institutions; and f) conduct a select number of high priority pilot projects in health care, business infrastructure development, and state government functions.

## **Engineering Laboratories**

### *Document Imaging Laboratory*

This laboratory supports DocView, MARS and other research and design projects involving document imaging. Housed in this laboratory are advanced systems to electro-optically capture the digital images of documents, and subsystems to perform image enhancement, segmentation, compression, OCR and storage on high density magnetic and optical disk media. The laboratory also includes high-end Pentium-class workstations running under Windows 98 and NT, all connected by 100 Mb/s Ethernet, for performing document image processing. Both inhouse developed and commercial systems are integrated and configured to serve as laboratory testbeds to support research into automated document delivery, document archiving, and techniques for image enhancement, manipulation, portrait vs. landscape mode detection, skew detection, segmentation, compression for high density storage and high speed transmission, omnifont text recognition, and related areas.

### *MARS Production Facility*

Designed, developed and maintained by Center staff, this facility houses high-end Pentium workstations and servers that constitute MARS-1 and MARS-2 production systems used by operators to scan medical journals and also keyboard data to produce MEDLINE records. While primarily for production, this facility also serves as a laboratory to collect data for the continual improvement of the MARS systems, such as a large collection of bitmapped document images, zoned images, labeled zones, and corresponding OCR output data. This collection serves

as a test set for research into techniques for autozoning, autolabeling, autoreformatting, intelligent spellcheck and other key elements of MARS.

### *Image Processing Laboratory*

The Image Processing Lab is equipped with a variety of high-end servers, workstations and storage devices connected by 100 Mb/s Ethernet. The laboratory supports the investigation of image processing techniques for both grayscale and color biomedical imagery at high resolution. In addition to computer and communications resources and image processing equipment to capture, process, transmit and display such high-resolution digital images, the laboratory also has a variety of image content. Most machines are equipped with multiple networking ports (FDDI, ATM, Ethernet, fast Ethernet) which allow, in addition to standard networking capabilities on the local Ethernet, the capability of alternate physical communications channels with these machines. This capability has been used in communications engineering experiments for point-to-point satellite channels connecting these machines with remote sites. ATM switches connect the Ethernet and FDDI networks to other local area networks throughout the building, to the Internet, and to experimental ATM networks such as ATDnet and MCI's research network, in addition to vBNS, the infrastructure for the Next Generation Internet and Internet-2 initiatives.

## **External Research Support**

### *Telemedicine*

NLM's Telemedicine Program is designed: (1) to evaluate the impact of advanced networking on health care, research, and public health; (2) to test methods to preserve the privacy of individual health data while also providing efficient access for legitimate health care, research, and public health purposes; and (3) to assess the utility of emerging health data standards in health applications of advanced communications and computing technologies.

NLM is the lead agency within DHHS for the government's High Performance Computing and Communications initiative and as such has a direct interest in the use and effects of advanced networking on health care. The growth of the Internet and the increasing access to high-speed computers and communications by consumers, health care providers, public health professionals, and basic, clinical, and health services researchers is having a fundamental

effect on health and human services throughout the nation. Major research and evaluation issues included in our telemedicine program arising from the current and future impact of advanced networking include: (1) the impact of telemedicine on the health care system as a whole and on cost, quality, and access to care for specific populations; (2) the benefits of integrated access to practice guidelines, expert systems, bibliographic databases, electronic publications, and other knowledge-based information from within computer-based patient record systems and other automated systems that support research and practice; (3) the maintenance of patient confidentiality as increasing amounts of electronic health data are transmitted via telecommunications during health care and aggregated for important public health and research purposes; and (4) the development of data standards and uniform practices for effective transmission, aggregation, and integration of health care, public health, and research data.

Nineteen multi-year telemedicine projects, affecting rural, inner-city, and suburban areas, with a total budget of \$42 million are currently under way. The projects, located in 13 states and the District of Columbia, will serve as models for: evaluating the impact of telemedicine on cost, quality, and access to health care; assessing various approaches to ensuring the confidentiality of health data transmitted via electronic networks; and testing emerging health data standards. Each project has reviewed and is applying the recommendations from two NLM-sponsored National Academy of Sciences studies: one on criteria for the evaluation of telemedicine (*Telemedicine: A Guide to Assessing Telecommunications for Health Care*); and the other on best practices for ensuring the confidentiality of electronic health data (*For the Record: Protecting Electronic Health Information*).

#### *Next Generation Internet*

A three phase effort to support test-bed projects that demonstrate the use of Next Generation Internet capabilities by the health care community are also under way. The projects are to be designed to improve our understanding of the impact of NGI capabilities on the nation's health care, health education, and health research systems in such areas as cost, quality, usability, efficacy and security. Phase 1, which began at the end of FY 1998, was a 9-month planning effort. Planning contracts were awarded to 24 institutions to identify the relevant outcomes, processes and cost variables and present a strategy for their measurement. During FY 1999 a 3- year Phase 2 effort

was started. Fifteen research contracts were awarded to support the implementation of Phase 1 plans within a limited geographic scope. Phase 3, a 2-year effort projected to begin in FY 2002, will test the scalability of Phase 2 projects to a national scope.

NLM is sponsoring a study by the Computer Science and Technology Board of the National Research Council to define the technical capabilities that the NGI must provide in order for it to support the demands of health care applications. The study will identify likely health care applications; examine their demands for such characteristics as bandwidth, quality of service, security, and access; and recommend an appropriate strategy for implementing these capabilities within the NGI. A special effort will be made to distinguish those capabilities that are unique to health care applications from those required by more general NGI based applications. The results of this study are scheduled for release during FY 2000.

#### *Visible Human*

During FY 1998, NLM and the National Institute for Dental and Craniofacial Research (NIDCR) jointly sponsored a workshop on the feasibility of creating a multimedia head and neck atlas based on Visible Human data. Based on the recommendations of that workshop, NLM and NIDCR have joined with the National Institute for Deafness and other Communication Disorders, the National Eye Institute, and the National Science Foundation to sponsor several parallel efforts. One project will develop an interactive head and neck atlas based on the current Visible Human data. The atlas will reside on computers housed at NLM and will be made available via the Internet. The atlas will have labels, voice pronunciation of anatomical terms, and other visual and explanatory information. A modular approach will be employed so that new advances in technology and segmentation techniques can be rapidly inserted as the project develops. We envision that the atlas will be immediately upgraded as new enhancements to various components are developed and added.

During FY 1999 a two year contract to bring the Visible Human Atlas of the Head and Neck to reality was awarded to Dr. Victor Spitzer of the University of Colorado Medical Center in Denver. A second research project will deal with the "computational" problems of segmenting and aligning Visible Human and patient specific CT and MRI data. The research effort will focus on developing new computational methods and packaging these methods as a public domain software toolkit with a common API

and data structure. During FY 1999 six 3-year contracts were awarded in order to form a research consortium to work on this problem. A third research project will be designed to resolve the "anatomy" problems and to overcome the deficiencies in the current data caused by the anatomical data acquisition method that was used. In general, finer resolution of all the organs and structures is desired. Specific research problems include finding new ways to prevent brain herniation as an

artifact of freezing the specimens; improving imaging of structures related to the eye; establishing appropriate techniques for arterial injection to maintain appropriate contrast; developing an appropriate technique for staining nerves during cryosectioning; and imaging nerves and other structures at a very fine level of detail. Proposals will be solicited and contracts to form a research consortium to work on these problems will be awarded during FY 2000.

# NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION

*David Lipman, M.D.*  
*Director*

The National Center for Biotechnology Information (NCBI) was established by Public Law 100-607 in November, 1988, as a division of the National Library of Medicine. The establishment of the NCBI reflected the importance of information science and computer technology in the understanding of the molecular processes that control health and disease. NCBI celebrated its 10th anniversary in November 1998, marking a decade of growth for a center recognized as a national resource for molecular biology information.

From its inception, NCBI was charged with providing data access and analysis tools for molecular biology information. Over the past 10 years the management of biological information has truly come of age, becoming increasingly integrated into the scientific process. It is now almost impossible to think of an experimental strategy in biomedicine that does not involve some online foray into scientific databases. At the core of this shift is a huge data explosion, most notably in the amount of gene sequence and mapping information. As its 10th anniversary year draws to a close, the horizon is a familiar one—a flood of data coming in many new forms. Through the next decade, NCBI will continue to keep pace with the flood of genome data, while also designing the tools and databases for the gene discoveries of the 21st century.

The Center has been given the responsibility to:

- Create automated systems for storing and analyzing knowledge about molecular biology, biochemistry, and genetics;
- Perform research into advanced methods of computer-based information processing for analyzing the structure and function of biologically important molecules and compounds;
- Facilitate the use of databases and software by biotechnology researchers and medical care personnel; and,
- Coordinate efforts to gather biotechnology information worldwide.

NCBI consists of senior scientists,

postdoctoral fellows, and support staff working in computational biology research and the application of these methods to building public information resources. These scientists have backgrounds in medicine, molecular biology, biochemistry, genetics, biophysics, structural biology, computer and information science, and mathematics.

NCBI programs are divided into three areas: (1) creation and distribution of sequence databases, primarily GenBank; (2) basic research in computational molecular biology; and, (3) dissemination and support of molecular biology databases, software, and services. Within each of these areas, NCBI has established a network of national and international collaborations and also closely coordinates its activities with other NLM divisions.

## **GenBank: The NIH Sequence Database**

NCBI is responsible for all phases of GenBank production, support, and distribution. GenBank is the NIH genetic sequence database, an international database that collects all known DNA sequences, and a critical research tool in the analysis and discovery of gene function.

The growth rate of GenBank continues exponentially. In FY 1999, over 2,000,000 new sequences were added. The two-million mark in terms of number of sequences was reached in January 1998, the three-million mark in December 1998, the four-million mark in June 1999, and five-million sequences are expected to be reached in December 1999. Likewise, the first billion basepairs were accumulated over a 17-year period, the second billion basepair mark in 14 months, the third in eight months, and the fourth will be reached in only five months. For the coming year, the rate of growth shows no signs of abating. The continued support for sequencing of the human genome and refinements in automated sequencing technology promise that the exponential growth rate will only increase.

NCBI produces GenBank from thousands of sequence records submitted directly from authors prior to publication; records submitted to our international collaborators, EMBL (European Molecular Biology Laboratory) at Hinxton Hall, UK and DDBJ (DNA Data Bank of Japan) at Mishima, are shared through an automated system of daily updates. Other cooperative arrangements, such as with the U.S. Patent & Trademark Office for sequences from issued patents, augment the data collection efforts and ensure the comprehensiveness of the database.

GenBank indexers with specialized training in

molecular biology create the GenBank records and apply rigorous quality control procedures to the data. NCBI taxonomists consult on taxonomic issues, and, as a final step, senior NCBI scientists review the records for accuracy of biological information. Improving the biological accuracy of submitted data and correcting existing entries are high priorities for the GenBank team. New releases of GenBank are made every two months; daily updates are made available via the Internet and the World Wide Web.

Comprehensive coverage of all sequence data, protein as well as DNA, is provided by GenBank along with the corresponding MEDLINE bibliographic information, including abstracts and links to publishers' full-text documents. New this year were links to textbooks and outside sources for obtaining the full-text paper when there is no link to the publisher. The latter service, called LinkOut, also points to other resources, such as biological databases and sequencing centers. GenBank is a key component in an integrated sequence database system that NCBI developed in order to serve as a single, comprehensive source of all known DNA and protein sequence information. The integrated database offers researchers the capability to perform seamless searching across all available data including the MEDLINE abstracts linked to the sequence data.

GenBank has evolved to contain several types of DNA sequences, from relatively short Expressed Sequence Tags (ESTs) to assembled genomic sequences several hundred kilobases in length. EST data obtained through cDNA sequencing are critical to understanding gene function and continue to be heavily represented in GenBank. As such, additional annotation is available for these sequences as part of the separate EST database (dbEST). During FY 1999, dbEST increased by 62%, totaling approximately 3,000,000 ESTs. The database continues to expand with the addition of 400,000 human ESTs from the Cancer Genome Anatomy Project, which arrive at approximately 10,000 ESTs per week. This is in addition to the ESTs that are generated by the Merck/Washington University project for humans and other organisms.

Another rapidly increasing segment of GenBank is the GSS (Genome Survey Sequences) division, which increased by over 160% in FY 1999. These sequences are similar in nature to ESTs, except for their genomic origins, and additional data on each sequence is kept in the separate Genome Survey Sequences database (dbGSS).

The STS (Sequence Tagged Site) records division of GenBank experienced significant growth in the past year. Sequence tagged sites are short sequences

that are operationally unique in the genome and used to generate mapping reagents. A separate Sequence Tagged Sites database (dbSTS) contains additional sequence data, such as information about the contributors, experimental conditions, and genetic map locations. During FY 1999, dbSTS increased by 38% and expanded to approximately 85,000 publicly available sequences.

The High-Throughput Genomic sequences (HTG) division of GenBank was created to accommodate a growing need to make "unfinished" genomic sequence data rapidly available to the scientific community. Incomplete sequences (phase 1 and 2) are updated in the HTG division as work progresses, and moved to the relevant organismal GenBank division upon completion (phase 3). The division experienced enormous growth in FY 1999, increasing by over 300% in number of sequences. HTG sequences are substantially longer than sequences in other divisions of GenBank, averaging about 114,000 basepairs, and the 91% increase in basepairs during FY 1999 reflects the impact of the work of the Human Genome Sequencing Centers worldwide.

Complete genome sequences for *Escherichia coli* K-12 MG1655, *Plasmodium falciparum* chromosomes 2 and 3, *Chlamydia pneumoniae*, *Helicobacter pylori* J99, *Thermotoga maritima*, *Pyrococcus abyssi*, *Rickettsia prowazekii* strain Madrid E, and *Aeropyrum pernix* K1 were added to the Entrez Genomes database. Other organisms for which there was a 50 percent or more increase in the number of GenBank records are *Oryza sativa* (rice), *Drosophila melanogaster* (fruit fly), and *Mus musculus* (house mouse), *Rattus norvegicus* (Norway rat). Sequences from these organisms continue to provide valuable clues for understanding the functioning of human genes. There was an increase of 64 percent in human sequences, which reflects the continuing accelerated pace of sequencing centers worldwide that are participating in the Human Genome Project.

### **The Human Genome**

NCBI has assumed responsibility for collecting, managing, and analyzing the growing body of human genomic data generated from the sequencing and genome mapping initiatives of the Human Genome Project, responsibilities previously held by the recently dissolved Genome Data Bank. NCBI has collaborations with large sequencing projects that are producing complete genome records, and plays a key role in assembling and making public the smaller records that can be linked and integrated with the Entrez Genomes

database.

As the Human Genome Project nears completion, the research focus is turning from analysis of specific genes or regions to a whole genome approach. NCBI has developed a suite of genomics resources to support comprehensive analysis of the human genome. Specialized tools and databases have been designed to facilitate the use of this data.

Completion of the sequencing and analysis of the human genome promises to be a complex task that will involve cooperation among researchers applying diverse tools to the problem. The human genome data generated will be reported in a variety of forms reflecting this diverse set of tools. Genetic and physical maps, markers, nucleotide polymorphisms, disease phenotypes, expression profiles, and sequence data must be integrated and made accessible for analysis. A repository of sequence data represents a natural site for construction of a nexus into which data in many forms can flow and from which these data can be accessed. NCBI's new Human Genome Resources Web page, designed to serve as such a nexus, is closely connected to the GenBank sequence database and also provides centralized access to a full range of human genome resources available within NCBI and elsewhere. Genome maps, sequencing progress, data on genetic variation and gene expression are included on the site, along with access to genome resources, such as GeneMap '99 and Genes and Disease.

The Human Genome Sequencing site is also an important new resource supporting the human sequencing effort. The site provides current information on genome sequencing progress and access to data. Chromosome-specific BLAST searches, text queries, browsing by chromosome, and sequence downloading are available, as well as access to constructed genomic contigs and international sequencing center data.

Of particular interest to the scientific and academic communities, as well as to the lay public, is the Gene Map of the Human Genome, which supplements a paper in the October 23, 1998 issue of Science. Produced in collaboration with a team of 64 scientists worldwide, the map charts the chromosomal locations of over 30,000 human genes, about twice the number of the earlier version of the map, which was released in October 1996. NCBI's UniGene (Unique Human Gene Sequence Collection) was used in the compilation of the Gene Map, and specifically, more than 75,000 clusters in UniGene represent more than 75% of all human genes. The Gene Map will greatly expedite the discovery of human disease genes and is expected to result in advances in detection and treatment of common illnesses. To date, the physical

map has assisted directly in identifying about 100 disease-causing genes. For scientists, the Gene Map Web page contains background information on STS (sequence-tagged site) markers, RH (radiation hybrid) mapping, and provides marker and map views, as well as options to do text or position searching. The Gene Map site also links to the Genes and Disease Web page, designed for the public and students. It includes descriptions for some 60 genetic diseases and links to databases and organizations that provide additional resources. A new Coffee Break Web page provides a collection of short reports on recent biological discoveries. Each report incorporates interactive tutorials that show how bioinformatics tools are used as a part of the research process.

The new Reference Sequence (RefSeq) database provides curated sequence data and related information for the community to use as a standard. RefSeq standards provide a foundation for functional annotation of the genome, offering a stable reference point for mutation analysis, studies of gene expression, and polymorphism discovery. In contrast to GenBank, which is a repository of sequences, RefSeq database is a non-redundant set of reference sequences, including constructed genomic contigs, mRNAs, and proteins. In the future, it will contain entire chromosomes.

LocusLink, a single-query interface to curated sequence and descriptive information about genes, was also launched this past year. LocusLink presents information on official nomenclature, aliases, sequence accession numbers, phenotypes, EC numbers, MIM numbers, UniGene clusters, map information, and relevant Web resources. LocusLink serves as an intercommunication tool for staff at NCBI, Human Genome Organization (HUGO), and Online Mendelian Inheritance in Man (OMIM).

NCBI continues the public distribution of OMIM, an electronic version of Dr. Victor McKusick's catalog of human genes and genetic disorders. OMIM has passed the 10,000 record mark and usage exceeds 5,500 users per day.

Another database of importance to the participants in the Human Genome Project, called dbSNP, was initiated in late FY 1998, and in its first year of public availability, dbSNP has received nearly 17,000 submissions. A database of single nucleotide polymorphisms (SNPs), dbSNP also contains both the experimental conditions used to find each mutation and each mutation's observed variation for populations and individuals. The information in dbSNP will be useful for physical mapping, disease association, and surveys of population structure.

The new SAGEMap database of quantitative

gene expression was introduced in FY 1999. Serial Analysis of Gene Expression (SAGE) refers to a technique for taking a snapshot of the messenger RNA population of a cell. By using the SAGE technique, coupled with high-throughput sequencing technology, it is possible to obtain accurate expression data for thousands of genes within a cell. A major application of SAGE is in the identification of abnormal gene expression leading to, or diagnostic of, various disease states, such as cancer.

The Cancer Genome Anatomy Project (CGAP), a collaborative effort with the National Cancer Institute, brings together data on gene expression in normal, precancerous, and malignant tissues. The CGAP database currently contains expression data for over 20,000 human genes. During FY 1999, the CGAP Web site was augmented with several new resources, including Gene Express, a new utility to find the computed expression level of genes in all EST libraries; the Mitelman Chromosomal Aberration Summary, a genome-wide map of chromosomal breakpoints in human cancer; and the Cancer Chromosomal Aberration Project (CCAP), designed to expedite the definition and detailed characterization of the distinct chromosomal alterations that are associated with malignant transformation. In addition, the SKY (Spectral Karyotyping) database, which contains chromosomal structures in cancer, is in development.

### **PubMed**

Late in FY 1999, a beta version of a completely redesigned PubMed was released that incorporates many capabilities requested by the medical librarian community who will be using PubMed in place of the mainframe search software that was retired. Functions were added or improved for limiting queries by common search filters, such as language, review, and human; building queries from alphabetic term lists; combining search statements by number; saving search results in a temporary clipboard; and displaying records. Context-specific help and a "frequently asked questions" section provide guidance in making the transition to the new system.

In addition, two major changes to the links from PubMed were made. A component called LinkOut was added to provide a more efficient way to store and manage the external links from PubMed and the other Entrez databases. It also broadens the PubMed linking function so that any content provider—publisher, data aggregator, library, specialized research database—can contribute custom URLs that link from specific PubMed records to their own Web sites. A second type

of new link from PubMed records is called, simply, Books. Links to relevant sections of the textbook, *Molecular Biology of the Cell*, have been made from words in the PubMed abstract and index terms. Another text, *Retroviruses*, will be added soon, and topical areas under consideration for future implementation include immunobiology, emergency medicine, and a medical dictionary.

Participation of the publishing community continues to increase, and the number of links to full-text journals increased by two-thirds in 1999. At the end of FY 1999, there were links to nearly 500 full-text journal sites. Some publishers provide access only by subscription, and others provide free access to the current issue, with subscription required for back issues.

PubMed usage had extraordinary growth, virtually doubling over FY 1998. PubMed has the highest growth rate of NCBI's services: each day over 700,000 searches of PubMed are performed. Usage of PubMed by scientific and lay populations, encouraged by free access since June 1997, grows each month, with over 16 million searches per month in late FY 1999. As many as 100,000 different users seek information daily via PubMed. Data on users and usage was collected during a one-week survey via PubMed; the largest user groups were researchers and physicians, and the leading uses of the data were academic research, education, and patient care.

In collaboration with the NIH Director's Office, NCBI is playing a key role in establishing a Web-based repository for barrier-free access to primary reports in the life sciences. This repository will be called PubMed Central, based on its natural integration with the existing PubMed biomedical literature database. PubMed itself will extend its coverage of the life sciences and continue its linkage to external online journals. PubMed Central will archive, organize, and distribute peer-reviewed reports from journals, as well as reports that have been screened but not formally peer-reviewed. The core concept of the proposal is to remove the barriers of access to the scientific literature and to make it available worldwide to the scholarly community at the least cost. NCBI has been involved in the development of this project, and has hired staff to develop the necessary resources for the storage and dissemination of the electronic journal articles.

### **Other Specialized Databases and Tools**

Comparison, whether of morphological features or protein sequences, lies at the heart of biology. The introduction of BLAST in 1990 made it easier to rapidly scan huge sequence databases for overt

homologies and statistically evaluate the resulting matches. With more than 8,700 citations to date, the paper describing the original algorithm has since become the most heavily cited of the decade. BLAST compares a user's unknown sequence against the database of all known sequences to determine likely matches. Sequence similarities found by BLAST have been critical in several gene discoveries. Hundreds of major sequencing centers and research institutions around the country use this software to transmit a query sequence from their local computer over the Internet to a BLAST server at the NCBI. In a few seconds, the BLAST server compares the user's sequence with up to a million known sequences and determines the closest matches.

Not all significant homologies are overt, however. Some of the most interesting are subtle and do not rise to statistical significance during a standard BLAST search. NCBI has extended BLAST and its statistical methodology to address the problem of detecting weak but significant sequence similarities and developed Position-Specific Iterated BLAST (PSI-BLAST), which searches sequence databases with a profile constructed from BLAST alignments. For protein analysis, the new Pattern Hit Initiated BLAST, or PHI-BLAST, complemented the profile-based searching that was previously introduced with PSI-BLAST. PHI-BLAST incorporates hypotheses as to biological function of a query sequence and restricts the analysis to a set of protein sequences that are already known to contain a specific pattern or motif. Another new feature added to the BLAST suite of programs is Organism-specific BLAST, which allows users to limit a BLAST query to a specific organism or taxonomic group. A utility to retrieve parts of sequence for unfinished genomes was also added. The BLAST Web documentation was augmented with instructional papers on interpreting BLAST statistics.

The BLAST sequence searching server is one of NCBI's most heavily used services and its usage continues to grow at a pace reflecting the growth of GenBank. Each day more than 70,000 sequence searches are performed, with users submitting their requests through e-mail, server/client programs, and the World Wide Web. The e-mail service has a public key encryption option to guarantee the confidentiality of user data as it traverses the public networks. The popularity of BLAST has stressed the existing computing capacity and additional computing resources have been added to accommodate the growing volume of users and expansion of the sequence databases. A new system, QBLAST, was developed to better handle the increasing BLAST load. This system obviates the

need for persistent connections while users are waiting for results and allows NCBI to better distribute the load.

Several specialized Web pages were released or substantially redesigned in FY 1999. The COGs (Clusters of Orthologous Groups) database, a natural system of gene families from complete genomes, was further enhanced. Analysis of COGs shows the molecular similarities and differences between species, which not only can provide clues about evolution, but also may help to identify protein families, predict new protein functions, and point to potential drug targets in pathogenic species. COGs now incorporates 21 complete genomes, tripling the number of organisms initially represented. In addition to phylogenetic patterns, the COGs may now be searched using free-text words or protein and gene names.

Collaborations with other NIH Institutes for disease-based services continue to flourish. Examples include the development of Web sites for analysis of genetic diversity in malaria and HIV. The Malaria Genetics and Genomics Web site, developed in collaboration with the National Institute of Allergy and Infectious Diseases (NIAID), contains *Plasmodium falciparum* sequences databases, genome maps, and linkage markers. This redesigned site now includes expanded BLAST databases, computed annotation of chromosome 3, and a rodent malaria division. In support of retrovirus research, NCBI launched the HIV-1 Subtyping Tool in FY 1999. This new Web-based tool was developed to facilitate monitoring of HIV-1 genomes.

The members of the NCBI taxonomy group plan the overall structure of the taxonomy database and Web pages, monitor the literature, and maintain contact with off-site taxonomy advisors and the collaborating sequence databases in Europe and Japan. They add new species or perform other edits to the taxonomy database when so required by new systematic insights and guide the NCBI indexing staff on taxonomic issues. The taxonomy group is currently composed of four biologists and one computer scientist. A Metazoan taxonomist with a primary specialty in the fish will begin in early FY 2000.

The taxonomy database contains 65,508 organisms represented with one or more sequences in GenBank (42,600 at the species level). About 900 species a month were added to the database in the past year, for an overall increase of 42%. Improvements were made to the taxonomy browser in the past year, including a bulk mover for easier modifications to the classification schema, links to the structure and genome databases. The taxonomic database is recognized as the

standard reference by the international sequence database collaboration. During FY 1999, the NCBI taxonomists provided consultation to both EMBL Data Library and DNA Database of Japan staff and, with the Structure Group, added taxonomic information to PDB-derived records.

## **Database Access**

### *Entrez Retrieval System*

The major database retrieval system at NCBI, Entrez, was originally developed for searching nucleotide and protein sequence databases and related MEDLINE citations, and has been expanded to include the integrated set of PubMed, MMDB (Molecular Modeling Database) 3-D Structure, Genomes, and Taxonomy databases. Users can search gigabytes of sequence and literature data with techniques that are fast and easy to use. A key feature of the system is the concept of "neighboring," which permits a user to locate related references or sequences by asking for all papers or sequences that resemble a given paper or sequence. The ability to traverse the literature and molecular sequences via neighbors and links provides a very powerful yet intuitive way of accessing the data. Entrez's design permits incorporating additional linked databases without changes in the user interface. Web Entrez now provides graphical views of nucleotide and protein sequences and access to the NCBI Genomes database, which contains graphical views of sequences and chromosome maps. The structure viewer, Cn3D, permits visualization of 3-dimensional protein structures. The latest version offers a greatly expanded array of annotation tools, including the ability to define molecular features and specify their display characteristics, global save capabilities, and an improved installation process. MMDB, the Entrez 3D Structure database, surpassed the 10,000-record mark in FY 1999.

A major revision of Entrez is under development to accommodate additional data resources and facilitate more finely tuned searches demanded by the explosion of data in the sequence, structure, and literature databases. This new release will have a completely new back-end, more databases, a new advanced search screen that records a history of all queries done, and a clipboard to save articles of interest for later use.

Entrez users submit over 500,000 text searches and 70,000 sequence similarity searches daily. Over 150,000 Entrez DNA and protein queries per weekday are handled and the number continues to

increase. Over 130,000 users access NCBI per day. Of these users, some 20,000 have accessed exclusively the molecular biology databases. Because of the mission-critical nature of NCBI's computing platforms for PubMed, Entrez, BLAST, and other services, an extensive program in system monitoring has been implemented. Based on measurements taken every 15 minutes from 50 sites across the U.S. and overseas, the average time to load the entire NCBI home page is under 2 seconds, and overall availability has been better than 99 percent.

### *Other Network Services*

Usage of NCBI's Web services, first introduced in December 1993, continues to expand as more services are added. NCBI staff continued to make access and usage easier with improved documentation and tutorials. Information about NCBI, its databases and services, data submission and update, and individual scientists' research projects are readily available, as well as an ever-increasing number of search tools. The Web server provides capabilities for Entrez and BLAST searches and submission by BankIt. Many other Web servers have links to the NCBI server to conduct searches and obtain the latest GenBank records. At the end of FY 1999, NCBI's site was averaging over 7,000,000 hits daily.

GenBank is also distributed over the Internet through the standard File Transfer Protocol (FTP) program, and many large commercial and academic sites maintain a local copy of GenBank. NCBI's Data Repository, with over 50 additional molecular biology databases, is also distributed via FTP; over 60 gigabytes of data, including daily GenBank and dbEST update files, are downloaded daily by over 1,000 sites. There are 2,000 FTP requests per day, increasing to 5,000 at the time of the bimonthly GenBank release.

NCBI maintains two electronic mail servers, BLAST and QUERY. The BLAST server performs sequence similarity searches and QUERY retrieves records from several sequence databases, including GenBank, EMBL, Swiss-Prot, and PIR. Any user in the world with e-mail access can submit a query to the servers and have an answer returned within minutes. More than 11,000 queries are handled daily by the BLAST and QUERY servers.

The improvement of NCBI's sequence submission software continued to be a high priority. Two new versions of Sequin, NCBI's stand-alone submission tool, were released in FY 1999. Sequin now allows the author to edit complete bacterial chromosomes or large eukaryotic chromosomal

segments in a single record. In addition, Sequin can now function as a stand-alone or network-aware program. Complete documentation and a tutorial are maintained on the Web site. The Web submission tool, BankIt, now in its fifth year of use, is still heavily used, but with the growing use of Sequin, submissions entering via this route have dropped from about 50% to just over 40% in the final quarter of FY 1999.

During FY 1999, NCBI upgraded a number of its key systems to keep pace with the increase in demand for public services, such as BLAST and PubMed, as well as to accommodate the dramatic increase in the growth of GenBank. Two clustered SGI Origin servers that act as gatekeepers to the NCBI's public Web site were upgraded from two CPUs each to four. One new 20-CPU SGI Challenge XL servers and nine 4-CPU Dell servers were added to support sequence similarity search services. Two of the three primary Sun Enterprise servers that support PubMed and Entrez were upgraded to faster CPUs and the number of those CPUs was increased from six to eight, and one gigabyte of additional memory was also added to each system.

In addition, approximately six major servers that support NCBI research and development activities were upgraded to newer models or newer and faster CPUs. One key database server received approximately 500 GB of additional disk storage. To keep up with the additional demands placed on the backup system by increased disk storage, a robotic tape library was installed that accommodates approximately four terabytes of off-line storage.

Equally important as building databases for molecular sequence information is the ability to access and retrieve the information using automated systems. The NCBI software toolkit concept addresses this need by creating software modules that provide a set of high-level functions to assist developers in building application software. Among these tools are a Portable Core Library of functions in the C language that facilitate writing software for different hardware platforms and operating systems, and AsnLib, a collection of routines for handling ASN.1 data and developing ASN.1 applications. The ASN.1 (Abstract Syntax Notation) tool is an International Standards Organization data description language that provides a mechanism for defining and structuring data as well as a set of program definitions that can interact with databases structured in ASN.1. With ASN.1 definitions and the NCBI software toolkit, complicated analysis programs can be readily constructed from pre-existing sets of modular tools, saving considerable time and programming effort.

## Basic Research

Basic research is at the core of NCBI's mission. The Computational Biology and Information Engineering Branches at the NCBI are comprised of a multidisciplinary group of scientists who carry out research on fundamental biomedical questions at the molecular level by developing and utilizing mathematical, statistical and other computational methods. The approach is both theoretical and applied. These two lines of research are mutually reinforcing and complementary. The basic research has led to new practical methods and the application of these methods has opened new areas of research.

There have been a variety of applied and theoretical studies of biologically important molecules and their functions, as well as continued development and improvement of algorithms and statistics for their analysis. Computer program and algorithm development have included the development of protein sequence search methods using patterns as seeds; a pairwise sequence comparison method; database design, data management and analysis for single nucleotide polymorphism (SNP) data; measures of threading statistics; and an analysis of the interactions among quantitative traits in the course of sympatric speciation.

Macromolecular sequence analysis programs have been applied to investigate chaperone-like ATPases; POZ domains; HD domains; "death" domains in apoptosis; DNA polymerase beta-like in nucleotidyltransferases; GPS domains in polycystin-1; START domains in signaling proteins; selenium-containing proteins; HIV-1 haplotypes and HTLV-I subtypes; and AT-hook motif containing proteins. In addition, many other families of proteins from previously published works are being further studied in the view of identifying new members, especially in fully sequenced genomes.

Genome-scale projects have continued to be a staple of Computational Biology Branch research. Projects have included genome analyses using clusters of orthologous groups of proteins sequences from whole proteomes; the distribution of protein folds in the three superkingdoms of life; the comparison of complete proteomes of *Caenorhabditis elegans* and *Saccharomyces cerevisiae*; the evolution of protein families in Archaea; finding systematic annotation errors in whole genome annotations; the use of metabolic pathways for functional annotation of genome sequences; the analysis of regulatory sequence elements in cell cycle-dependent transcription; the

analysis of the genome of the human pathogen, *Chlamydia trachomatis*; and the analysis of the genome of the malaria parasite, *Plasmodium falciparum*.

Other efforts have included the further development of a database of histones and histone fold sequences and structures; further research on a model of the HIV infectivity; a genetic linkage map of microsatellites in the domestic cat; analyses of the horizontal transfer of genes using the evolution of aminoacyl-tRNA synthetases as an example; and the prediction of curved DNA in promoter sequences.

The intramural group is engaged in over 20 projects, many of which involve collaborations with NIH and other research laboratories. The work is reviewed by a Board of Scientific Counselors of extramural scientists (see Appendix 4 for list of members). The high caliber of the work has been evidenced by the number of peer-reviewed publications, over 100 in FY 1999, and the requests for outside collaborations. The staff has made 76 presentations and mounted 14 posters at major scientific meetings, and made 44 presentations to academic departments and companies engaged in molecular biology research. In addition, they made 25 presentations at NCBI's computational biology lecture series and hosted presentations by 48 invited speakers.

The Visitors' Program continues to be successful in bringing members of the scientific community to the NCBI to engage in collaborative research with the Computational Biology Branch as well as joint activities in database design and implementation with the Information Engineering Branch. This program, administered in conjunction with Oak Ridge Associated Universities, facilitated 71 visits by 65 individual senior researchers, including 3 long-term sabbaticals, this past year.

In order to provide for concentrated efforts on improving and strengthening GenBank, the GenBank Fellow post-doctoral program at NCBI to provide for concentrated efforts on improving and strengthening GenBank; all positions are filled, using the NIH IRTA (Intramural Research Training Award) Program Fellowships and the Fogarty Visiting Fellow mechanism to recruit for this program.

### **User Support and Outreach**

As part of its mandate to facilitate the use of databases and software by the biology community, NCBI maintains a user support group staffed by scientists, librarians, and information specialists with broad experience in handling biology and medical information. The primary focus of this group is to

support the particular services that NCBI offers by e-mail, phone, and fax. NCBI has extended its outreach to the library science community by invited presentations and workshops on biotechnology information topics.

As the number of database services and number of users has increased, the scope of user support services has also expanded. NCBI staff in the Information Resources Branch, with contractor support, provide responses to queries for information and assistance. The three main areas of user support include: information about GenBank and related molecular biology database services and data submission; technical assistance for submission of new GenBank data and revision of existing records; and technical assistance with Entrez and other data retrieval services. Most responses are immediate and nearly all answers or information are provided within 24 hours of receipt of a message. Authors who submit their sequences to GenBank are furnished with accession numbers for publication within 48 hours. During FY 1999, assistance with general PubMed inquiries from both users and publishers was provided by staff from MEDLARS Management and Reference Sections, NLM.

To increase awareness of NCBI and its programs, NCBI staff participate in exhibits, seminars, workshops, and courses, both nationally and internationally. In FY 1999, NCBI staffed exhibits at scientific society meetings, including the American Society for Human Genetics; American Society for Cell Biology; American Association for Cancer Research; The Protein Society; American Society for Microbiology; Special Libraries Association; American Society for Biochemistry and Molecular Biology; Association of Biomolecular Resource Facilities (ABRF); and Experimental Biology (FASEB). Additional exhibits were staffed at more specialized meetings, including TIGR/Computational Genomics; TIGR/Microbial Genomes; Human Genome Meeting (HUGO); Understanding Genetics Through Information Technology; Computational Genomics Conference; and NIH Research Festival. Under an expanded training initiative, User Support staff taught eight courses in the use of NCBI's databases and services to scientists, librarians, and information specialists at six sites. In addition, shorter training lectures were given at four sites.

NCBI members participated as faculty at one or more courses sponsored by the American Association for Cancer Research; The Johns Hopkins University; George Mason University's Institute for Computational Sciences and Informatics; the Jackson Laboratory; Marine Biological Laboratory; Cold Spring

Harbor Laboratory; International Center for Genetic Engineering & Biotechnology in Trieste, Italy; Experimental Biology '99; Northwestern University; and City of Hope National Medical Center.

NCBI staff also participated in courses taught on the NIH campus, including a lecture/workshop on NCBI services for over 80 NIH staff, and two courses in linkage analysis, sponsored by the Center for Information Technology. NCBI also co-sponsored, with NHGRI, a mini-course in genomics that attracted over 100 attendees at each of a series of lectures.

The NCBI participates in an advisory and collaborative role with other government agencies such as the Patent and Trademark Office and the Department of Agriculture on programs involving biotechnology information. Within the NIH, the NCBI coordinates with other institutes and particularly with the National Human Genome Research Institute and the National Cancer Institute on databases and informatics programs that impact information exchange on a national level. The NCBI newsletter was distributed to a mailing list of over 23,500 biologists and institutions—this was a slight decrease from FY 1998 due to the introduction of PDF-formatted issues on the Web. New fact sheets and tutorials on programs and services were distributed at all public forums where NCBI was represented.

NCBI's Web site was completely redesigned and restructured to accommodate the growth of all services, with particular emphasis on the addition of databases and tools used in human genome research. The new design not only offers the needed flexibility and organization, but also provides added interest through frequently updated announcements and news items.

### *Extramural Programs*

Funding for extramural bioinformatics activities is the responsibility of another division of the NLM, the Extramural Programs. It offers a program of grants for computer analysis of molecular biology data. A wide variety of work in computational biology has been supported through the program, including methods and algorithms for sequence analysis, structure and function prediction, new machine architectures and specialized databases. Extramural postdoctoral training in the cross-disciplinary areas of biology, medicine, and computer science is also funded through the NLM's informatics fellowship program.

### *Biotechnology Information in the Future*

The explosive growth in the fields of genetics and molecular biology and the application of this knowledge to medical practice reinforce the need to build and maintain a strong infrastructure of information support. NCBI continues to be engaged in developing and employing new methods for disseminating knowledge to the biomedical community. Based on a core of advanced intramural research in several areas of computational biology, NCBI rapidly addresses the evolving informatics needs for genome research with state-of-the-art software and databases. Genomic information resources such as NCBI have repeatedly demonstrated their value as indispensable discovery tools for basic research. The value of these resources will only continue to grow as they support the breakthroughs in basic research and provide us with better understanding and treatment of human disease.

# EXTRAMURAL PROGRAMS

*Milton Corn, M.D.*  
*Associate Director*

The Extramural Programs Division (EP) continues to receive its budget under two different authorizing acts: the Medical Library Assistance Act (unique to NLM), and Public Health Law 301 (covers all of NIH). The funds are expended as grants-in-aid, and in some instances as contracts, to the extramural community in support of the goals of the NLM. Review and award procedures conform to NIH policies.

EP issues grants in several categories for which the motif in general is medical informatics:

- Resources for Information Management
- Training
- Research and Research Resources
- Publication
- Educational Technology
- SBIR/STTR
- Other

## Resource Grants

Resource Grants, authorized by the Medical Library Assistance Act, support access to information as well as promote networking, integrating, and connecting computer and communications systems. There are four types of Resource Grants which range in complexity relative to their intent as well as dollar amounts and duration of time. They are considered "seeding" grants designed to initiate a resource or service or program, and to help such to eventually become self-sustaining. The commonalities among the four Resource Grants are that public and private, nonprofit health institutions/organizations engaged in health education, research, patient care, and administration are eligible and that there must be library involvement.

### *Information Access Grants*

Information Access Grants, aimed primarily at hospitals, clinics, community health centers and similar small health institutions and organizations, support computers to access NLM's PubMed and other Internet resources; it also funds the automation of public access library catalogs. These grants provide up to \$12,000 per institution participating and are available to single as

well as multiple institutions working together.

### *Information Systems Grants*

Information Systems Grants, ranging up to \$150,000 per year for up to three years, are aimed at larger hospitals, medical schools and institutions promoting the access and use of health information. These grants support computer access to information on a more complex level than Information Access Grants.

### *Internet Connections Grants*

The Internet Connection Grants provide funding for single institutions up to \$30,000 and multi-institutions up to \$50,000 to initiate Internet access for health institutions by funding gateway/router equipment, Internet Service Provider fees, and line charges in the first year. Some institutions with existing Internet access extend it to other institutions.

### *IAIMS Grants*

Integrated Advanced Information Management Systems (IAIMS) Grants are designed to support institution-wide information systems that link a variety of individual and institutional databases and information systems for patient care, education, research, library, and administration. IAIMS grants support two phases, planning and implementation, with the program goal being to support organizational mechanisms that foster the integration and sharing of various information systems. The planning phase funds up to \$150,000 for one to two years; the operational phase up to \$500,000 per year for five years or \$550,000 with an IAIMS apprenticeship option.

## Training and Fellowships (MLAA)

### *Training in Informatics*

Exploiting the potential of computers and telecommunication for health care information requires investigators who understand fundamental problems of knowledge representation, decision support, and human-computer interface. NLM remains the principal support nationally for research training in the fields of medical informatics, including clinical and basic science domains. NLM provides three mechanisms of support for its training activities.

Five-year institutional training grants support approximately 150 trainees at predoctoral and postdoctoral levels. Twelve institutions currently

receive such support. For the past few years, NCI and NIDR contributed funds to NLM to help support slots at these training sites for applicants interested in radiation therapy and dental informatics respectively. Following some staffing changes, NCI discontinued its support in FY 1999.

Individual informatics research fellowships are available to those who seek research training similar to offerings at the institutional training sites but at a site of their choosing. Individual applied informatics fellowships are available to individuals who want to learn informatics techniques and technology for application in their current professional specialties.

#### *Education of Health Sciences Librarians in Informatics*

All NLM informatics training programs have been encouraged to develop and offer training within the curriculum suitable for those interested in health science libraries. NLM agrees to provide additional funding for any slots awarded to librarians. Response has been gratifying and is growing. Librarians are now in place or soon to begin at: University of Pittsburgh, University of North Carolina, Oregon Health Sciences University, and University of Missouri.

#### *Training of Minorities*

NLM is participating in an NIH-Wide Fellowship Program aimed at encouraging under-represented minorities in research careers. In FY1999:

- Dr Carlton R. Moore, at Mount Sinai Medical Center in New York received an applied medical informatics fellowship for work on a database of free medications for indigent patients. His mentor is Dr Joseph Kannry.
- Fanny C. Hawkins at Baylor College of Medicine received an applied medical informatics fellowship for work on implementing and evaluating electronic medical records in ambulant care situations.

#### *Other Minority Support*

Internet Connection Grants were awarded to the Lassen Indian Health Clinic (Susanville, CA), the Pleasant Point (Indian) Health Clinic (Perry, ME), and the Yavapai College (Prescott, AZ). Information Access Grants were awarded to Charles Drew University in Los Angeles, Presentation College (Aberdeen, SD) for the Lakota campus whose students are from the Cheyenne River Sioux Reservation, and to the University of Arizona for three Pinal County institutions that serve significant minority populations.

An Information Systems Grant to UCLA will identify special collections of American traditional medicine from ethnic populations, particularly African American, Hispanic and Native American. Another Information Systems Grant to the New York Public Library funds information access for consumers in its branch libraries, some of which serve minority populations.

Research grants are made through a variety of mechanisms, including individual research projects, cooperative agreements, research resource grants and others.

#### *Medical Informatics*

NLM's research grants sponsor investigation of basic and applied medical-knowledge issues that arise at the intersection of biomedicine, computer science, and human behavioral sciences. NLM has been increasingly successful in recent years in interesting other NIH Institutions in supporting informatics projects wholly or in partnership with NLM.

#### *Biotechnology Informatics (Computational Molecular Biology)*

The techniques of informatics are indispensable tools for handling the complex data generated by molecular biology research. NLM continues to provide research grants for informatics projects in this area of basic medical science, as well as training grants, and grants for support of research resources. A related problem concerns the development and maintenance of electronic databases on which researchers increasingly rely, and for which no other source of support has yet been identified.

Because tools for sequence analysis are now well developed and widely available, EP is expanding the computational biology grant program into related areas including molecular evolution, population studies, and other research areas opened up by the availability of sequence data, including neuroinformatics, and modeling.

Following the June 1999 publication of the Biomedical Information Science Technology Initiative (BISTI) on biomedical computing, as requested by the Director, NIH, NLM together with a number of other Institutes began a continuing series of discussions about the various ways in which NIH intends to address national needs for training and research in biomedical computing. No new programs were issued in FY 1999.

NLM also participates with 15 other organizations in the Human Brain Project, which seeks

innovative methods for discovering and managing increasingly complex information in the neurosciences. Each participant selects grants within the project for full or shared funding.

### **Other Grants**

#### *Publication Grant Program*

The Publication Grant Program provides short-term financial support for selected not-for-profit, biomedical scientific publications. Studies prepared or published under this NLM program include critical reviews or research monographs in the history of medicine and life sciences; on special areas of biomedical research and practice; on medical informatics, health information science and biotechnology information; and in certain instances, secondary literature tools and scientifically significant symposia. Resources in recent years have been used principally for history of medicine projects standard hardcopy has been the most common format, but projects in electronic publishing, video, and other media were also supported. The program has an informal self-imposed ceiling of \$35,000 on direct costs per grant per year.

#### *SBIR/STTR (PHS 301)*

All NIH research grant programs, including NLM's, by Congressional mandate allocate a fixed percentage of available funds every year to Small Business Innovation Research (SBIR) grants. These projects may involve a Phase I grant for product design, and a Phase II grant for testing and prototyping. NLM also participates in the other mandated fund allocation program, Small Business Technology Transfer, but generally it contributes its small allocation to other NIH institutes, as it did this year.

#### *Conference Grants*

Support for conference and workshops is intended to help scientific communities identify research needs, share results, and prepare for productive new work.

#### *Biomedical Ethics*

Ethical issues in health care and research produce an enormous literature. This literature comes from law, medicine, public health, and government. The National Reference Center for Bioethics Literature at Georgetown University continues to offer invaluable

resources and guidance for workers in this area. An NLM contract maintains the Center. A complementary contract from Library Operations supports an indexing activity that contributes to BIOETHICSLINE, one of NLM's online databases.

### **Other Extramural Programs Activities**

In addition to its standing grant programs, Extramural Programs Division of NLM engages in a number of special projects aimed toward important biomedical goals, and often involving cooperation with another NIH institute or other Federal agency. Some examples of such activities in FY 1999 follow.

#### *The Digital Libraries Initiative—Phase 2 (DLI-2)*

This initiative explores innovative digital libraries research and applications. The program extends the previously sponsored "Research on Digital Libraries Initiative." The term "digital libraries" is used to denote the vast distributed collections of text and images available through the Internet. Much research and development will be needed before these new electronic libraries can be used easily and efficiently to obtain reliable information. DLI-2 is administered by the National Science Foundation and is jointly sponsored by the NSF, the Defense Advanced Research Projects Agency, the NLM, the Library of Congress, the National Aeronautics and Space Administration, the National Endowment for the Humanities, and others.

The project is interested in electronic information in a broad spectrum of fields in arts and science. Improving network-based information access for health care consumers is an important goal of the project for NLM, although all aspects of digital libraries as applied to health domains may compete for funding. NLM, as have the other sponsors, contributed funds to NSF, which manages the project. NLM's commitment for FY 1999 was \$1,000,000, and represents an arm of the HPCC initiative. Total project budget from all sources may exceed \$50 million over 5 years. NLM made available to interested applicants the Unified Medical Language System Knowledge Sources and the Visible Human datasets. Applicants were also free to use resources of their own choosing.

#### *Improving Public Health and Health Services Research*

Because of the remarkable potential of information technology to process huge quantities of data, there is growing interest by both professionals and the public in measuring the quality and effectiveness of

current medical practices. Noting that the Agency for Health Care Policy and Research (AHCPR) recently funded a number of training sites intended to increase expertise in health services research, NLM initiated a series of joint discussions and planning for a workshop to explore ways and means of using informatics, and in particular, the NLM informatics research training programs, to facilitate the AHCPR training goals, and to further our mutual interest in health services research and public health research. A jointly sponsored symposium on the topic will be held in FY 2000 at NLM.

#### *Informatics for the National Heart Attack Alert Program*

This program receives approximately two thirds of its funding from National Heart, Lung, and Blood Institute, and the remainder from NLM. In June 1999 a second Request for Proposals was published to obtain research and development using more innovative, promising, emerging medical informatics technologies as an approach to reducing or eliminating some or all of the obstacles hindering the ability of the NHAAP to reach its goals. In contrast to the FY 1998 solicitation, which focused on main-line informatics approaches, the 1999 RFP sought high risk, high payoff, high technology approaches. Five Phase 1 contracts for nine-month planning phases were awarded. Technologies to be explored include wearable devices, portable computing devices, and wireless communications devices. It is expected that subsequent phases will include prototype development and implementation contracts.

#### *Miscellaneous Special Projects*

NLM continues to transfer funds to other agencies to support projects of broad scope and utility for biomedical research. The agencies that received funds were the National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS), the National Institute on Deafness and Other Communication Disorders (NIDCD), and the National Institute of Mental Health (NIMH).

NLM received co-funding from other organizations, including Department of the Army, NIMH, and others.

#### **Grants Management Highlights**

The Grants Management staff reviews NLM grant applications for administrative content and

compliance with guidelines and directives; prepares and disseminates grant award documents; maintains official grant files for NLM; provides consultation and assistance to grantees on appropriate business management concepts; and advises NLM officials on grants management policy and procedures.

The Grants Management staff, which consists of four employees, issued a total of 146 awards for FY 1999. Effective October 1, 1998, the Notice of Grant Award for all of NIH was redesigned from a document format into a letter format. The same information is contained in the letter format. This change has enabled the NIH to e-mail the Notice of Grant Awards to the Grantee Institution instead of mailing hard copies.

#### *Review Committee Activities*

NLM's scientific merit peer review group, the Biomedical Library Review Committee (BLRC) met three times in 1999 and reviewed 128 applications of which 98 were approved. The Committee (see Appendix 5 for roster of members) operates as a "flexible" review group, i.e., it is composed of three standing subcommittees: eight members on the Medical Library Resource Subcommittee, nine members on the Medical Informatics Subcommittee; and four members on the Biomedical Information Subcommittee. The subcommittees consider research applications in medical library projects, medical informatics, and biotechnology information respectively.

A final peer review of applications is performed by the Board of Regents, which meets three times a year, approximately three months after the Biomedical Library Review Committee. (A roster of the Board members is in Appendix 2.) One of the Board's subcommittees, the Extramural Programs Subcommittee, meets the day before the full Board for the review of "special" grant applications. Examples of "specials" include applications for which the recommended amount of financial support is larger than some predetermined amount, when at least two members of the scientific merit review group dissented from the majority, when a policy issue is identified, and when an application is from a foreign institution. The Extramural Programs Subcommittee makes recommendations to the full Board, which votes on the applications.

Nine Special Emphasis Panels (SEP's) were also coordinated. They reviewed 59 applications, of which 58 received priority scores. Such panels are convened on a one-time ad hoc basis to review applications for which the regularly constituted review groups lack sufficient expertise. Use of SEP increased

significantly in FY 1999 because of new regulations requiring NLM to supervise SEP review of review of contracts as well as of grants.

#### *Review Reform*

The NIH is participating in the President's Reinventing Government initiative. A portion of this activity has been the establishment of a Peer Review Oversight Group charged with coordinating, evaluating, and making policy recommendations for all peer review conducted at NIH. The Group is continuing to meet and will consider input from the applicant community. Some changes which have occurred include:

- The Division of Research Grants has become the Center for Scientific Review, and a number of the standing study sections have been reorganized.
- Five specific criteria have been chosen as a basis for assessing the merit of NIH research projects grants.
- CFIRST (R29) awards for newly independent biomedical investigators became R01 (traditional Research Project) awards June 1998.
- Applicants wishing to apply for a FIRST award will indicate this on the cover of the application.
- New guidelines went into effect October 1, 1998

for the inclusion of children in NIH funded research involving human subjects.

#### *Administrative and Personnel*

EP's Administrative Officer retired during the year, necessitating recruitment for a replacement. In addition EP recruited for a Program Assistant, and a Committee Management Assistant for FY 2000.

#### *Summary*

NLM's Extramural Programs, like almost all extramural grant divisions at NIH, regrets that not all applications of good quality can be funded, but the grants that can be made are furthering NLM goals in most key areas. The situation was particularly acute in FY 1999 because a larger than average demand on the available budget came from the budget requirements of existing grants. A number of excellent grants, particularly in informatics research were held over to FY 2000 in the hope that the new budget will permit funding.

**Table 9****Extramural Grant and Contract Program***(Dollars in thousands)*

Category	FY 1997		FY 1998		FY 1999	
	No.	\$	No.	\$	No.	\$
Resource projects	40	5,307	39	5,897	48	6,339
IAIMS	(8)	(3,040)	(9)	(3,690)	(11)	(3,611)
Access	(6)	(320)	(3)	(198)	(8)	(1,025)
Systems	(11)	(1,412)	(10)	(1,352)	(9)	(1,025)
Connections	(15)	(535)	(17)	(657)	(20)	(678)
Research	80	14,994	72	15,887	87	53,635
Medical Informatics Projects	(32)	(5,955)	(38)	(8,384)	(43)	(12,000)
Medical Informatics Resource	(1)	(230)	...	.....	...	.....
Biotechnology	(27)	(6,259)	(16)	(4,730)	(6)	(2,972)
Cooperative Agreements	(2)	(754)	...	.....	...	.....
Career Awards	(18)	(1,796)	(17)	(1,773)	(11)	(1,199)
Library Science	...	.....	...	.....	...	.....
Digital Library	...	.....	(1)	(1,000)	(27)	(37,464)
Training	20	5,645	23	5,519	21	1,239
Institutional	(12)	(5,290)	(12)	(4,993)	(12)	(758)
Fellowship	(8)	(355)	(11)	(526)	(9)	(481)
Publications	9	247	8	269	5	222
Bioethics	1	498	1	513	1	528
SBIR/STTR	4	401	5	508	5	528
Regional Medical Library	8	6,611	8	6,710	8	8,706
NIH Tap		830		830		.....
<b>Totals:</b>	<b>162</b>	<b>\$34,533</b>	<b>156</b>	<b>\$36,133</b>	<b>175</b>	<b>\$71,197</b>

# OFFICE OF COMPUTER AND COMMUNICATIONS SYSTEMS

*Roy A. Standing*  
*Acting Director*

This fiscal year, OCCS simultaneously has been keeping legacy systems running while transitioning to support of a new re-engineered client/server web approach with an ILS and support for all of NLM. It is a substantial challenge for the Library to maintain transparency to tens of thousands of users while replacing systems central to its core functions. The complex transition from existing systems and services must occur in parallel with the development of new ones.

OCCS has been providing effective online access to NLM information for nearly a quarter of a century. The Library's computer systems, however, have relied heavily on 1970s technology, e.g., large mainframe computers and custom-built software. OCCS is aiding Library Operations and the rest of NLM to utilize the recent advances in software, hardware, and communications that now make it possible to provide new services to users. The goal of OCCS is to become a service-oriented applications provider and facilities manager for NLM.

The mission of OCCS is to provide efficient, cost-effective computing and networking services, technical advice, and collaboration in informational sciences in support of the research and management programs offered through NLM. OCCS develops and provides the NLM backbone computer networking facilities, and supports, guides, and assists other NLM components in local area networking. The Division provides professional programming services and computational and data processing facilities to meet NLM program needs; operates and maintains the NLM Computer Center; designs and develops software; and provides extensive customer support, training courses and seminars, and documentation for computer and network users.

OCCS helps to coordinate, integrate, and standardize the vast array of computer services available throughout all of the organizations constituting NLM. The Division also serves as a technological resource for other parts of NLM and for other Federal organizations with biomedical, statistical, and administrative computing needs. The Division

promotes the application of High Performance Computing and Communication to biomedical problems, including image processing. The OCCS staff develops computer-based systems for information retrieval applications, conducts computer science and engineering research and development, and consults and collaborates in the area of advanced electronic office automation facilities. They support software systems to perform these services, and conduct research and evaluations for best fit solutions to information access needs.

## **Year 2000 Concerns**

The Year 2000 (Y2K) problem was an enormous challenge. NLM has met this challenge and helped its users address their Y2K problems. As reported to the Board of Regents in January 1999, NLM is on schedule to be fully Y2K compliant. The Y2K problem has three components: 1) proper leap year calculation (2000 is a leap year), 2) coding the year with only two digits, and 3) co-mingling status codes in date fields. OCCS has reviewed several common ways that applications can be patched or fixed. The Y2K impact on NLM falls in four areas: mainframe systems, Unix systems, desktop systems, and LANs. OCCS coordinated the Library's approach to dealing with the problems in each of these areas. In previous fiscal years, Y2K solutions were provided to all licensees of NLM's databases, and Y2K solutions were developed for ELHILL, MEDLARS and DOCLINE.

*Several Y2K initiatives were undertaken during 1999:*

Independent Validation and Verification—OMB required that "mission-critical" systems receive independent certification by March 31, 1999. NLM identified three mission-critical systems: PubMed, Toxicology Data Network (TOXNET), and MEDLARS Database Updating (MEDU). All three systems underwent Independent Validation and Verification (IV&V) this year. This took several months, including running the applications in a "time machine" mode. All three systems were certified as Y2K compliant.

Y2K Remediation of Non-Mission Critical Systems—OCCS performed a survey of all NLM's IT application systems for inclusion in the NIH Year 2000 Inventory System. Sixteen major NLM systems were identified and added to the Inventory System. For all the systems with Y2K compliance issues, remediation and testing were successfully completed during 1999. All NLM systems have been reported to CIT and HHS as being Y2K compliant.

Y2K Day One Rehearsal—NIH has implemented a Year 2000 Day One Rehearsal plan for all of NIH. OCCS has been the point of contact for NLM's involvement in this plan, as well as the actual Day One Plan for New Year's Eve weekend. OCCS submitted a Day One plan for the IBM Mainframe System Support, LAN and Communications Systems, and Facilities Management Section. Desktop Review—OCCS coordinated the review of NLM desktops for Y2K compliance and remediation, bringing together the plans of NLM divisions. The identification and audit phase required teams of OCCS staff and contractors months to complete. Of the 1539 systems recorded in the ClickNet database, some 120 systems failed hardware Y2K tests. All but 5 five can be fixed; noncompliant systems have been upgraded or replaced. The ClickNet product also was used to inventory software titles installed on NLM desktops. 152 software titles were classified as "global" and 170 as "local." Many of these titles were already compliant. Remediation plans for the noncompliant global titles have been established, and updates of several (NT, Windows 95, MS Office) have been undertaken. As of fiscal year-end, desktop Y2K activities are proceeding with updates on targeted software and ongoing updates for other applications being installed. As of September, mission-related desktop software is 91% compliant.

### Development Branch

This has been a year of transition for the Development Branch of OCCS, which developed and now supports many of NLM's current products and services. As part of NLM's Systems Reinvention Project, OCCS developers also are at the forefront of NLM's transition to more modern systems. Though several legacy systems were retired in 1999, OCCS's most significant task remains that of ensuring that the legacy systems continue to provide uninterrupted service. These remaining legacy systems required substantial maintenance this year. This support also will be required in 2000. Still, the Branch this year began developing new applications that will replace the remaining legacy systems. OCCS thus is well positioned for a future of continuing support of the major mission requirements of NLM. This year OCCS implemented these new applications:

- VOYAGER Integrated Library System and LOCATOR*plus*, replacing legacy systems
- Two major releases of MEDLINE*plus*, NLM's consumer health Web site
- MeSH2000, which replaced a legacy system
- The redesigned NLM home page.

The following new applications currently are being developed by OCCS. Some will replace the remaining legacy systems, others expand application support to other divisions of NLM:

- Data creation and maintenance project—supports creation of MEDLINE and other databases
- DOCLINE—supports Interlibrary Loan Service, including SERHOLD and DOCUSER
- Publications project—supports production of *Index Medicus* and other paper publications
- LSTRC project—supports the Literature Search Technical Review Committee
- NLM Classification project—supports production of Pager and Web version
- External Data Distribution Project—supports exchange of bibliographic information
- OAMS Inventory Control System
- OAM Small Purchases Management System
- OAM Contract Management System.

Along with other divisions, OCCS contributed to the prestigious Hammer Award for NLM's System Reinvention efforts, which was awarded September 29 at a meeting of the Library's Board of Regents. OCCS accomplishments cited included Voyager, the new commercial integrated library system (ILS), and LOCATOR*plus*. Under NLM's system reinvention effort, a sophisticated integrated library system replaced an assortment of custom-built, mainframe-based processes that control acquisitions, serials, cataloging, collection management, circulation, preservation, and binding.

The OCCS Development Branch has two sections: The Information Management Section directly manages software analysis, development, testing, and support of NLM information access facilities, and maintenance to NLM applications. As the legacy systems age, the number of staff with the skills necessary to maintain these critical systems has decreased. A small team from this section is responsible for maintaining the legacy systems for as long as necessary while transitioning to new skills.

The Software Support Section provides needs analysis, project coordination, system design, and software development in support of NLM initiatives. The primary focus of this section is support of NLM Web services such as the NLM home page and MEDLINE*plus*. Staff also support NLM System Reinvention Projects.

*VOYAGER Integrated Library System (Phase 1)*

As a Federal Reinvention Laboratory under the National Performance Review, NLM is transitioning from its mainframe legacy systems to a more modern client/server environment. A key milestone in this transition was November 30, 1998, when the Voyager Integrated Library System (ILS) was placed into production. Implementation began with the release of Voyager for in-house use by NLM staff for cataloging and acquisitions work. Voyager supports NLM's internal support systems, one of the three parts of NLM's System Reinvention Project, the other two being access and retrieval.

Voyager is a fully integrated system that combines an open system architecture with relational database technology. It allows NLM to exchange data between the ILS and other unique applications in a straight-forward manner. Voyager manages acquisition and receipt of monographic and serial materials, cataloging of bibliographic and authority data, and circulation of items to individual patrons and interlibrary loans. It also provides access to LOCATOR*plus*, NLM's Online Public Access Catalog, via the World Wide Web. For many NLM program areas, Voyager will be the primary or only system needed.

To complete implementation, NLM staff worked with Endeavor, the vendor, to add additional features to the Voyager ILS. Key additions were enhancements to serials processing and to the closed stack request module. Extracting and converting information from NLM's disparate legacy systems into a single integrated system was extremely challenging. OCCS developed the software to export data such as patron information from DOCLINE into Voyager. Software was also developed to export of data in the US/MARC format to other institutions. Over 100 NLM staff members, mainly from LO and OCCS, worked together to make this implementation a success.

Voyager is now running on UNIX/Solaris servers with an underlying Oracle RDBMS. OCCS provides on-call support to maintain the availability of the servers and databases as well as to perform backup and restoration functions. OCCS reformats data from Voyager daily. It is available to outside users via FTP on a semiweekly, weekly, or monthly basis.

#### *Voyager Integrated Library System (Phase 2)*

The second phase of the Voyager ILS was placed into production on April 12, 1999. During this phase, OCCS installed the 1) Voyager Online Public Access Catalog (OPAC), renamed LOCATOR*plus* by NLM, and 2) Voyager Circulation module, which

supports the four reading rooms at NLM. LOCATOR*plus* replaced NLM's telnet-based legacy online card catalog Locator, which is no longer available.

LOCATOR*plus* provides access via the Internet to monographs, journals, reports and audiovisuals that are on order or in process at the library. LOCATOR*plus* has other new features. Anyone using the catalog from the web is able to:

- View the receipt of both serial and monograph materials;
- Find out if NLM owns a particular journal title;
- Search for books, audiovisuals and journals, either together or separately;
- Search limited collections, such as the history of medicine or reference collections;
- Search on field-specific areas using advanced search features;
- Obtain direct access to a variety of other resources, such as MEDLINE*plus*, Images from the History of Medicine, TOXNET, HSTAT, and other U.S. medical library catalogs.

Several software programs and techniques were developed by OCCS to ensure that Voyager works well in the NLM reading room environment, including:

- Secure Workstations—Workstations in the reading rooms are initialized to perform only reading room activities. The environment of each workstation is "locked in" so that a patron cannot alter the configuration of the workstation.
- Closing Notification—This software displays a daily message to notify patrons that the reading rooms will be closing soon.
- Overnight Photocopy Service (OPS)—This ORACLE form-based application processes requests from Voyager. Every five minutes, output is formatted for the Relais database. By clicking on an icon on the Voyager OPAC screen, patrons are able to view, and then print OPS requests while still in the reading room. Using the OPS Request System, NLM patrons find it easy to keep track of their photocopy requests.
- Closed Stack Operation—The standard process in the delivered Voyager system for requesting and filling requests for material from closed stacks is very labor intensive. It requires numerous keystrokes, mouse clicks and bar code readings for a request. Given the high volume of such requests at NLM, OCCS proposed an automated process. OCCS streamlined the process using two commercial products, MacroExpress and Microsoft Word. As a result, processing requests from the closed stacks is now extremely efficient.

### *ELHILL Monographs to Voyager*

ELHILL is one of the Library's workhorse legacy systems that slowly is being phased out. Maintained for the last 25 years by OCCS, the 23 databases of ELHILL provide online access to information about a wide range of biomedical subjects. A team of LO and OCCS staff members have begun to define conversion specifications for programs to export monographic data from various ELHILL databases into Voyager. The processes and programs needed to convert these nonstandard monographic records to USMARC format are complex; they also will be unique for each ELHILL database. Software is being developed to convert data from five databases—HSTAR, HISTLINE, SPACELINE, BIOETHICS, and POPLINE—to the ILS by the middle of year 2000.

### *Main Web Server Hardware Upgrade*

With the increased use and significance of the NLM Web home page and related services, a more failsafe environment was needed. This year, OCCS UNIX system administrators installed (April 1999) a three-server cluster to support the NLM main web site, MEDLINE*plus*, and other reinvention applications when released. The OCCS web team and UNIX systems administrators worked closely together. To get the new server cluster up and running, the web team ported the Web site administration tools and the back-end scripts and processes from the old to the new environment. The old environment had one development and only two production servers, i.e., the main web server and the MEDLINE*plus* server. In addition, the new environment also has a load balancing router and three other servers, namely an administration server, a supporting server, and an NFS file server.

OCCS Web development team performed several difficult tasks to support this project, such as:

- Cataloging all web site administration tools, their supporting scripts, and a variety of scripts associated

with web site management and daily processing of web logs. This included the web search engine, the link checker/document expiration system, and the Listserv processing.

- Assessing how well current and proposed systems configurations support the functionality of the programs.
- Assigning localities for all tools and scripts.
- Modifying and testing all scripts and tools to operate in the new environment.
- Archiving unused scripts.

The main Web server hardware was upgraded successfully. NLM Web based applications are now available with improved system security.

### *NLM Home Page Design*

Members of OCCS web team worked on the redesign of NLM's home page. The layout of the new NLM home page has five major areas. NLM programs and services are presented in these easily understood categories. All NLM pages now have a clean and simple layout with consistent navigational and graphic headers.

The OCCS web team created mock-ups of possible new designs and provided technical support to the redesign committee. Coordinating closely with LO, the OCCS web team created HTML files for the committee's evaluation of possible Web pages, headers and footers. The web team created scripts to replace files from old NLM web pages that did not conform to the official template. OCCS web team members and the LO team created guidelines for NLM templates for departments wanting to design their own headers. Finally, the OCCS web team members and LO finished the redesign of the NLM web site, providing a consistent "look and feel" to the web site. The public was introduced to the new NLM Web site on July 28, 1999. Feedback on the design has been very positive.

## *MEDLINEplus*

MEDLINE*plus*, introduced in October 1998, is a gold mine of up-to-date, quality health care information. It links the public to reliable and authoritative information that has been reviewed by NLM's information experts. The development of MEDLINE*plus* was a joint effort of the Public Services Division (PSD) of LO and OCCS. OCCS provides the technical expertise; PSD is responsible for the content and management of the site. MEDLINE*plus* was awarded the October 1999 FOCUS AWARD from the National Cervical Cancer Coalition (NCCC). The citation noted, "These excellent pages are helpful in directing consumers to resources containing information that will help patients research their health questions."

MEDLINE*plus* operates on the 3-server Unix cluster environment that ensures high availability to the main NLM home page. A significant design feature of MEDLINE*plus* is that the contents reside in an off-line Oracle database and static HTML pages are generated from that database each night. During FY 1999, the team implemented 5 major releases of MEDLINE*plus*. Two of these releases introduced new functionality and improved usability of the site. Three releases enhanced the infrastructure or improved site management for content providers

## *International Loansome Doc*

The new NLM International Loansome Doc (ILD) web site was developed by OCCS, becoming operational in May 1999. ILD lists non-U.S. libraries that provide document delivery world-wide using DOCLINE. Since the early nineties, Loansome Doc has allowed individual MEDLINE users to participate in the interlibrary loan system by entering requests from their terminals. International users increasingly are accessing MEDLINE via PubMed or Internet Grateful Med. In response to this demand, the ILD web site provides information about participating DOCLINE libraries worldwide and links to them. In addition, the public can search for libraries by country.

The project contains two components, an NLM Administration function and the public Web site. NLM's Collection Access Section, PSD/LO, maintains the Oracle database holding the ILD information. This allows the web pages to be dynamically generated. As new libraries become affiliated with Loansome Doc, they can be added to the ILD Oracle database and made available to the public site.

## *NLM Classification Schedule*

The legacy system now used for maintaining and distributing NLM classification schedules and related information will be replaced by a more documented and maintainable web-based application. The approach begins with converting the current TESSCLASS database from the TESS system in M204 to an environment compatible with the Voyager ILS. The new NLM Classification Schedule System will: maintain NLM classification schedules and the related subject indexes; provide online classification information to catalogers at NLM; provide access via the Internet to medical library catalogers around the world; and become the basis for further printed versions of the classification schedule. The OCCS project team worked closely with Library Catalog staff, who have accepted the requirements and systems analysis. The OCCS project team is evaluating different search engines and DBMSs, and has started design of the database.

## *Impromptu*

OCCS has begun implementing Cognos Impromptu, a commercial Web reports software package, as an ad hoc report writer for NLM Oracle applications. Impromptu is now being used for reporting by CUSTQ, RELAIS and the Voyager ILS. In the future, Impromptu will be used by reinvention projects such as DOCLINE and the new MEDLINE data creation system.

## *OAMS Inventory Control*

The 10 year old dream of the NLM Office of Administrative Management Services to have a customized, online customer Ordering/Inventory Control System has now come true. OCCS completed development of this web-based system in September 1999. The OAMS office supplies inventory management process is now automated, allowing NLM users to order office supplies online. The detailed hardcopy requests that drove the old manual requisitioning process have been eliminated. Reported inventory levels will be adjusted automatically by the system when vendor/customer orders are filled and new stock is received. The system has two major components: Customer Features and Administrative Features. Menu-driven customer features include three easy search options (category, item description, and stock number) and automatic email notification of backorders, new stock items, and order confirmations. Administrative features include easy tracking of vendor

and customer orders. The system provides automatic notification when quantities fall below reorder points. Reports can be generated dynamically to detail current program area/customer expenditures, vendor and customer backorders, and stock received.

#### *Citation Data Creation and Maintenance System*

Good progress was made on the Citation Data Creation and Maintenance System this year. Design of the overall application was completed, and the main workflow control structure was developed. A key component is the editor, which will be used by NLM indexers to index journal issues more efficiently. A satisfactory prototype editor has been tested. A major undertaking by LO and OCCS was a detailed analysis to determine the best way to implement diacritics in the new system. The goal is to have all current MEDLINE diacritics supported.

#### *ORACLE Database Services*

OCCS this year has increased its in-house expertise in Oracle dramatically. The Oracle DBMS is being used on a number of projects at NLM such as the UMLS, Relais, the MeSH prototype project, CUSTQ, HELPQ and the Voyager ILS. OCCS responded to this growing demand for Oracle by obtaining an NLM-wide license and designing a systems architecture that optimizes the use of Oracle licenses. Because of this licensing arrangement and architecture, NLM's DBMS costs are significantly below those of Oracle's standard licensing arrangements.

OCCS staff investigate new Oracle releases, typically on prototype systems, to determine how best to manage them. For example, the new DCMS application was developed using the Oracle8i database. The Q-Diagnostic tool for Oracle databases was evaluated, then purchased to improve problem diagnosis. OCCS developed several other prototype systems this year to automate LO processes, often in conjunction with ILS implementation. During this transition year, OCCS staff have continued to provide Database Administration support for the Legacy systems that these Oracle databases are replacing. LO and OCCS staff have worked closely to provide continuing coverage as LO switched from cataloging and indexing in legacy systems to using the new Voyager ILS. These new Oracle databases have achieved consistently high availability. OCCS support in 1999 included bringing up many of these systems or upgrading them to version 8.0.5 of Oracle.

New projects underway using Oracle include the DOCLINE reinvention project, which includes SERHOLD and DOCUSER; the DCMS reinvention project to support creation of MEDLINE and other databases; publications reinvention project to support *Index Medicus* and other paper publications; the OAMS Inventory Control System; and the NLM Classification system.

#### *Web Page Promotion System*

Creating and deploying the NLM home page and related web pages should be performed in a controlled environment to ensure the integrity of the NLM Web site. OCCS will replace the current system, developed several years ago. The new design incorporates two environments. Contributors prepare their HTML pages in a test environment. After their contents are validated, web pages are promoted to a separate environment for the public site. Working with LO, the OCCS development team has begun to provide better functionality, security, and tracking in a more robust system. Different tools such as Coldfusion and Java servlets are being considered to supplement the current back-end Perl scripting. The new mechanism centers on a redesigned document tree with separate areas for HTML files and web application files.

#### *Medical Subject Headings (MeSH2000)*

Replacement of the mainframe-based software system for maintaining the Medical Subject Headings (MeSH) thesaurus (used for cataloging, indexing, and searching) is a major step in the NLM Systems Reinvention Project. The current MeSH legacy system was developed more than a decade ago. OCCS has completed development of a new custom system, Mesh2000. Beta testing is under way. Full production is scheduled for October 1999. As part of the reinvention project, the underlying data structure of MeSH is being altered to afford a concept-based representation of the terms and meanings in MeSH. The new concept-based structure is more similar to the structure of the Unified Medical Language System, which could make it easier to move the two systems into a common environment. At the same time, the role of entry vocabulary within a thesaurus and of the relationships between terms in a thesaurus have been clarified.

The MeSH2000 client/server architecture was implemented using current versions of Sun's Java DK and JFC. The client program runs on platforms that support this software, including Windows 95, Windows

NT, and Sun Solaris. The client exchanges data with the Oracle database server using stored procedures and SQL queries invoked through JDBC, the Java standard for database communication. The MeSH2000 server runs on a Sun Sparc Ultra-2 under a SunOS version of UNIX with an Oracle DBMS. The server portion of the software is implemented using Oracle data structures written in PL\*SQL.

#### *External Data Distribution Project*

Early in 1999, OCCS began an effort to provide data distribution in a more modern format than the magnetic tape required currently by NLM legacy systems. The new distribution approach will rely on Extensible Markup Language (XML) and File Transfer Protocol (FTP). This new distribution approach is possible because of the replacement of NLM's legacy systems. In 1999, however, NLM continued to distribute a large volume of databases and updates via magnetic tape. For example, 140 tapes are needed to hold all of the 20 databases of MEDLINE records for 1966 to 1998.

NLM is keenly aware that tape and tape cartridges are not state-of-the-art. They are cumbersome to produce and distribute. They are difficult for many of licensees to receive and process. The wide variety of NLM databases and varying frequency of tape shipments increases the complexity of NLM's data distribution program. A 1997 study recommended that NLM use a combination of FTP and 4mm DAT tapes both for distribution of weekly/monthly updates of MEDLINE and for yearly distribution of the MEDLINE regeneration file. DLT tapes will be used for distribution of large amounts of data, such as for complete MEDLINE shipments for new licensees. OCCS is approaching the final stages of transition to the new method of distribution. FTP will be used as the medium except for large amounts of data, which will be distributed on DLT tapes.

#### *DOCLINE*

As part of NLM's reinvention plan, DOCLINE, NLM's automated interlibrary loan request and referral system, will be moved to a new web-based platform in early 2000. DOCLINE provides electronic document delivery service to health professionals by rapidly routing interlibrary loan requests among libraries in the National Network of Libraries of Medicine. Since 1985, DOCLINE has been running on a mainframe computer at NLM while serving over 3,000 U.S. and Canadian medical libraries. It handles

more than 10,000 ILL transactions per day at no cost. DOCLINE access has also been granted to selected major national and medical libraries in other countries. The new system will combine DOCLINE with DOCUSER and SERHOLD and will retain all existing functions and incorporate many new features.

Work on the new DOCLINE this year has involved analysis, design and development of the web server and tools; the organization of Oracle Tables for DOCUSER, SERHOLD and the transactions; and the organization of Web Interfaces, i.e., PubMed, Loansome Doc, and ILS. The project is now in the third phase, usability testing. Feedback from NLM staff and DOCLINE participants is being evaluated. Deployment of the new web-based DOCLINE system is currently scheduled for the first quarter of calendar year 2000.

The DOCLINE development team has optimized performance of web operations such as page rewrites, database calls, and variable passing. The new system incorporates features that were not possible on the mainframe, for example it will display previous delivery information and support searching of PubMed and the ILS. The ISO interface will be available for rapid transfer of batches of preformatted requests. The system is being implemented using Coldfusion with standard SQL Oracle interfaces supplemented by JavaScript. The team used C++ for complex algorithms such as routing.

#### **Systems Technology Branch**

This has been a particularly challenging year for the Systems Technology Branch. This branch now supports a wide range of client/server hardware while still supporting the mainframe. It has expanded support of new Oracle applications throughout NLM, while simultaneously maintaining—and by necessity upgrading—some legacy systems and shutting down others. This year, OCCS renovated the computer room, then implemented more automated monitoring together with world class security. The branch is moving rapidly towards a facilities management approach that provides a safe, secure round the clock environment for NLM systems.

The Systems Technology Branch has four sections:

- System Services Section runs the NLM computer center with its associated telecommunications and large scale computer processors.
- Operations Section provides systems programming support to mainframe operating systems, NLM retrieval engines, and management support utilities.

- Network Engineering Section ensures proper network operation of a vast array of Novell IPX and Internet IP file and application servers.
- Desktop Services Section provides Library-wide software user assistance, administrative workstation support, and workstation hardware platform integration, upgrades, and configuration coordination.

#### *System Services Section*

The Mainframe Systems Team manages the large-scale computer processors and associated telecommunications equipment in the NLM computer center. The team had three goals for FY 1999:

- Preserving the past by maintaining NLM's mainframe legacy systems as they are phased out;
- Maintaining the present by shutting down old applications while simultaneously introducing new client/server applications; and
- Planning for the future by preparing the staff to support a very different systems environment after the mainframe is shut down.

Throughout FY 1999, NLM mainframe applications remained at a high state of availability and performance. Even though there were fewer applications and users on the mainframe, many support duties remained. Significant maintenance was performed on the IBM Mainframe Operating System, several dozen IBM and non-IBM software products, and hardware components such as the IBM 9672/R52 Enterprise Server, disk drive farm, communications controllers, and related hardware units. All commercial software products on the mainframe required upgrading during FY 1999. Procuring, deploying, testing, and migrating these upgrades to the production environment required months of work.

Several major difficulties were addressed this year in maintaining the current applications, including the increasing failure rates of the old disk farm, increasing security concerns, loss of contractor support for the ELHILL and DOCLINE applications, and Y2K concerns. Phased replacement of the entire disk farm required a great amount of time for research, procurement, and installation. Maintenance savings alone from this disk upgrade will recover the cost of purchase in 17 months. An added benefit is that this much more reliable hardware also consumes less power and uses less air-conditioning.

NLM faces an inherent need for increased computer system security as our users and NLM's exposure to the outside world through the Internet increases. The OCCS Security Project increased

protection of the 16,000 plus files on the mainframe and created numerous Alert/Warning mechanisms. The mainframe is now much more secure than ever in the past.

An unexpected loss of contractor support was a possible risk to ongoing ELHILL/DOCLINE operations. Through extensive study and careful planning, however, Systems Team staff members were able to prevent any loss of service. The team now operates these critical NLM mainframe applications without contractor support. Y2K compliant software was installed, tested, and put through Independent Verification and Validation. The team created a dedicated, isolated virtual test machine on the mainframe, which proved invaluable as a Time Machine to support Y2K testing. The close of the year found the Systems Team both confident and vigilant in terms of Y2K.

Looking toward a future in a decentralized client-server environment, mainframe staff began to take steps to augment mainframe skills acquired over many years. All the staff developed training plans and attended numerous classes on Unix, Oracle, Windows/NT, and Networking. The mainframe Systems Team had expected FY 1999 to be full of challenges and opportunities. It proved to be a year of accomplishment and productive change for both the team and for NLM's remaining legacy systems.

#### *Operations Section*

The Operations Section is moving toward a facilities management approach. This section provides systems programming support to mainframe operating systems, NLM retrieval engines, and management support utilities. Beginning in FY 1998, other divisions besides LO have been provided with floor space, power, system monitoring, system interaction and system reporting on a round the clock basis. NLM systems are supported in a safe, secure environment. This year, the section upgraded the facility significantly, providing better monitoring, more space, and improved power support.

In addition to the ongoing support of NLM/LO applications, OCCS operations have expanded the monitoring support provided to applications of other NLM divisions. This year, OCCS began to support PubMed for NCBI and TOXNET for SIS, particularly in order to provide after-hours support. Operations staff monitor these applications, using newly installed monitoring tools and alarms. For all NLM divisions, the section continues to provide testing support, including stress testing, for applications being

moved off the mainframe. Support for NLM security has increased. Audible alarms have been programmed to signal intrusions so that intruders can be trapped and damage to applications prevented.

Electrical power in the computer room has been upgraded and simplified. OCCS has been reorganizing the placement of equipment to free up power panels. Power panels are being consolidated so that problems can be isolated much better. As more UNIX servers are brought in, power lines supporting lower amperage are being installed. The computer room is now completely supported by Uninterrupted Power Supply (UPS). Particularly since the installation of UPS, unscheduled down time has become very rare.

The operations section has a service orientation, scheduling and meeting deadlines regularly. This is the only NLM organization that works around the clock 365 days a year. Staffing has remained level even while the section is supporting the move to UNIX client/server boxes while still having to maintain a number of legacy systems, including the laborious backing up and shipping of tapes.

#### *Network Engineering Section*

The Network Engineering Section (NES) this year continued to provide a better network infrastructure for NLM, building on the groundwork laid in FY 1998. Last year, NLM email systems were consolidated and the core LAN networking environment was upgraded. In FY 1999, the section was able to concentrate on improving network performance, management, features and services. Instead of constantly resolving operational problems, the section now focuses more on performance tuning for optimal throughput. NES operates an array of networks, including Novell IPX and Internet IP based file and application servers, and provides LAN support for network-related services for Novell, NT, and some Unix systems. This includes management of telecommunications and network services for NLM. The upgrade of the NLM LAN communications architecture from 10BaseT Ethernet to 100BaseT Ethernet using a switched technology was completed this year. NES is now finishing the documentation for the equipment (cabling, jacks and patch panels) upgraded and the wire management within each rack. Currently 4 of the 14 racks have wire management fully implemented. Implementation of web-based management of closet switches will continue during next year.

The Cisco 5500 Ethernet switching equipment installed during FY 1998 was upgraded this year.

Software was updated on both the 5500 RSM's and the 7513's IOS. The Supervisory III engines allows trunking between switches. This increases network performance significantly because traffic flows between 5500's without having to be routed. Work has begun to reconfigure the connections between LHC, NCBI, OCCS and CIT.

A number of system-monitoring packages were implemented at NLM during FY 1999 to improve management of the infrastructure including MailCentral (monitors the GroupWise mail system), Compaq Insight and Dell OpenManage (monitor the status of Compaq and Dell servers), and DS Expert and DS Analyze (monitor Novell NDS). Trend and analysis reports from these and other packages should enable further improvements in network performance. Implementation of HP Openview is the next step in overall network management and is one of the major NES goals for FY 2000.

The Novell environment was enhanced significantly during FY 1999. The Netware Directory Services (NDS) was upgraded to support new Netware 5 servers. The 5 new Netware servers are based on Compaq Proliant 6500 hardware. These enterprise level servers provide high levels of redundancy and protection for GroupWise post offices. The current NLM environment still includes Netware 4.11 and Netware 3.12 servers. Next year, all Netware 4.x systems will be upgraded to Netware 5. Once all systems are Netware 5, additional features such as Netware over IP will become available throughout NLM.

GroupWise services increased in performance and reliability this year. The hardware platform was upgraded, and GroupWise services were separated from login and printing services to provide greater performance. The GroupWise system was upgraded to version 5.5 on the servers and the clients to provide many new features, the most important being native addressing and auto spell check. When released, the version 5.5 client enhancement pack will add LDAP support at the client and speed up Web Access with Java Scripts. The beta version of the enhancement pack has been fully tested and found to be very solid. Limited installations have been done within NLM.

The NT Domain began taking on more importance within NLM during FY 1999. The existing PDC, BDC and Member servers were upgraded to Dell 2300 and 4300 servers. This not only improved performance but also provides a uniform hardware platform. IIS web services were implemented for a few OCCS applications such as Knowledge base, Attachmate web and Impromptu report writer. Further

work is planned to test IIS for greater use within the OCCS development area.

Production Internet connectivity for NLM continued to be provided through GTE/BBN Planet. T3 (45Mbps) connectivity to the GTE/BBN Planet network node in Washington, DC, is provided via a SONET ring. A T3 link for CIT/NIH to the node in Vienna, VA, is also provided. PSC and NCI links to the Internet are also provided. NLM and NIH collaborate in using these links to back up each other's Internet connectivity in the event that one link fails. The IT functions of the NLM Extramural Programs Office continued to be supported at EP's off-campus location in the Rockledge I building in Bethesda. On-site technical support was provided for the PC, network, and IMPAC II systems.

### *Information Systems Laboratory*

The OCCS Information Systems Laboratory has used the transition from legacy applications and the mainframe as an opportunity to build a new infrastructure for NLM. While NLM had only three small UNIX servers in 1992, OCCS now manages 62 UNIX servers. The DBA role for Oracle databases now performed by OCCS is supported by Oracle Enterprise Manager 2.0 and Q-Diagnostics. These tools have automated the monitoring and maintenance of NLM's Oracle databases.

This restructuring of the infrastructure in 1999 means that OCCS now provides much better performance for NLM applications systems. This improved performance is a result of the transition to better, faster UNIX boxes and the Oracle DBMS. The applications systems also have higher availability and redundancy because of the RAID disks. In addition, upgrading all subnet 74 network connections to the CISCO 5500 switch has enhanced both performance and network security. All ISL UNIX systems are now networked through the CISCO switch. This upgrade makes NLM transmissions more secure because only the address is passed along to the switch, whereas with a router the entire text is sent. Such enhanced security is needed because the Internet is inherently open and less secure. OCCS has implemented security at host level, intrusion detection, and security audits, including:

- Installing the Internet Security Systems' RealSecure intrusion detection software system to perform real-time monitoring of NLM's T3 Internet traffic. The Real Secure system monitors all inbound traffic and generates an alarm if it detects predetermined sets of circumstances keyed to suspicious activity such as possible hacker attacks.

- Installing the Lucent firewall, which is currently in a test mode.
- Using the Internet Security Systems' Internet Security Scanner to perform SATAN-like scans against NLM host systems to detect possible security vulnerabilities.
- Initiating in March an NLM-wide white-hat penetration analysis and security audit.
- Working with the FBI on serious incursions.

To support the new NLM home page, OCCS configured a high-availability WWW server environment consisting of a BigIP3 load balancing router, multiple Sun UNIX servers, and a Network Appliance NFS server. This environment houses the NLM home page along with MEDLINEplus, LOCATORplus and other NLM WWW applications. In building this new infrastructure, OCCS staff have working hard to be at the cutting edge of hardware technology. With the mainframe planned to shut down next year, the infrastructure for the future has now taken shape. The transition to a new client/server and UNIX/Windows NT environment based on state-of-the art Sun servers will be completed next year. The overall goals for OCCS in supporting this new infrastructure are to ensure high availability and security and to rely on a web front-end as the major avenue to NLM systems.

### *Desktop Services Section*

The Desktop Services Section exceeded its Service Level Agreement (SLA) goals this year and expanded its use of standards and of automated tools to provide ever better and faster service. The IT Services Center (formerly the OCCS HelpDesk) is open Monday-Friday from 7:30 A.M. to 5:00 P.M. It is staffed by 5 employees and 2 students.

During FY 1999, the Services Center handled approximately 5,000 HelpQ trouble tickets, 2,400 telephone calls, 275 e-mails, and 350 walk-ins. The average response time was 3 hours for High tickets, 2 days for Medium tickets and 5 days for Low tickets. This is less than half the response time required by the Service Level Agreements. Another responsibility is the configuration, installation and user verification of new and previously used workstations and peripherals. Approximately 500 workstations were installed for NLM users this year. The average build and install time per user was one half day. The Desktop Services contractors, Digicon Corporation and MIL Corporation, also played a substantial role in the Y2K audit of NLM's workstations and servers. Over 1,200 units were

tested in less than two months. Over 800 of NLM's workstations and servers were upgraded with the Y2K upgrade of Service Pack 4 (SP4) for Windows NT during nine days in January.

This year, the section taught 89 classes, with 822 total students, in Microsoft Office applications and GroupWise. A one-year pilot training partnership with the National Institute of General Medical Sciences began in September to evaluate alternative means of providing software training in the MS Office Suite. Responses from NIGMS employees have been very positive. If results continue to be favorable, NLM may enter into a more permanent arrangement. In addition, approximately 350 employees attended the six showings last winter of "The Plugged-In Mailbox: E-mail Uses and Misuses" video.

OCCS focuses on creating an environment in which deaf/hard of hearing and hearing people can communicate and work together effectively and productively. This year, NexTalk for Networks was implemented. It is a client communications software for use with NXi Test Services (NTS). NexTalk for Networks is designed for person-to-person "text telephone" use, providing compatibility with the TTY's used by the deaf and hard of hearing. It is also used for text conversations and exchange messages with other NexTalk users.

The Desktop Services Section played a key role in developing desktop standards and in increasing centralized buys, both of which will simplify desktop

support functions as well as lower costs. Two joint LO/OCCS committees have addressed user needs, customer service standards, and set standards for PC hardware and software systems. This has simplified PC acquisition thus reducing costs and saving time.

OCCS negotiated with providers of several standard software titles based on NLM's previously unrecognized eligibility for educational volume licensing. As a result, LO and OCCS have made joint purchases to re-license several software titles from Microsoft, Corel, and other software publishers. Such joint purchases saved over \$100,000 in software license costs last year.

Automated PC software deployment was used actively this year. The NAL (Novell Application Launcher) continued to save many hours of work. It has been used to deploy thousands of updates and applications. Other pilot projects, such as the migration to a 32-bit operating system, have been reviewed.

Since NAL can only deploy to workstations on the Novell environment, the next step is to collect workstation hardware and software inventory using ZenWorks workstation management tools. Such bundling and deployment tools continue to improve, providing an efficient way to maintain software at the desktop. OCCS will continue to evaluate such "Desktop Management" approaches for use in PC asset tracking, more efficient software distribution, and PC troubleshooting.

# ADMINISTRATION

Donald C. Poppe  
Associate Director for Administrative Management

## National Performance Review

The NLM System Reinvention is a high priority initiative conducted by NLM in support of its role as a reinvention laboratory under the National Performance Review. The project is designed to reinvent the Library's information systems, to move to a more flexible, powerful, and maintainable computer system that will improve internal processing and provide innovative services to outside users. Significant progress in system reinvention was made in several areas in FY 1999:

**Integrated Library System:** NLM acquired and installed Voyager, an integrated library system (ILS). The system supports library functions such as acquisition of books and serials, cataloging, serials control, collection management and circulation. It replaces several legacy systems that have been developed at NLM over the past 25 years. Voyager includes a Web-based online catalog, named LOCATOR*plus* at NLM, which replaces the earlier Locator system.

**PubMed Retrieval System:** Entrez, the software system underlying NLM's popular PubMed service underwent a major redesign in FY 1999. The new system, which became available as a test version in September, provides advanced searchers with new features such as search limits and a search history. It also integrates access to a number of databases at the National Center for Biotechnology Information, including the genome, nucleotide sequences, and protein sequences databases.

The MEDLINE database in PubMed was expanded to include the journal citations that have been in the HealthStar database. The use of PubMed continued to grow dramatically during the year. By September 1999, PubMed was averaging 95,000 users a day. The number of searches done each month went from 9 million in September 1998 to 18 million in September 1999. As another step in migrating from our legacy systems, on September 30, 1999, NLM discontinued access to the ELHILL retrieval system, which command language searchers have used for more than 25 years to search MEDLINE and other NLM databases.

**MeSH:** at the end of the fiscal year, NLM completed development of a new system to maintain

and manage NLM's database of Medical Subject Headings, MeSH. This further reduces dependence on legacy systems and software products.

**Other Developments:** In FY 1999, NLM also made significant progress in developing two other new systems: an improved DOCLINE and the Database Creation and Maintenance System (DCMS). The new version of DOCLINE, which serves NLM's interlibrary loan network, will be a Web-based system and will replace a system originally designed to work with TTY terminals. The DCMS, which will support the creation and maintenance of the MEDLINE database and auxiliary data collections, will also be Web-based and will replace numerous custom-built legacy systems

## Financial Resources

In FY 1999, the Library had a total appropriation of \$181,309. Table 10 displays the FY 1999 budget authority plus reimbursements from other agencies.

### TABLE 10

**Financial Resources and Allocations, FY 1999**  
(Dollars in Thousands)

<u>Budget Allocation:</u>	
Extramural Programs .....	\$ 39,833
Intramural Programs.....	133,315
Library Operations.....	(63,393)
Lister Hill National Center for Biomedical Communications.....	(43,391)
National Center for Biotechnology Information .....	(18,400)
Toxicology Information.....	(8,131)
Research Management and Support .....	8,161
Total Appropriation .....	\$181,309
Plus: Reimbursements .....	8,602
<b>Total Resources .....</b>	<b>\$189,911</b>

The 1999 appropriation language authorized the Library to use personal services contracts and provided for the availability of \$4.0 million without fiscal year limitations. These authorities are key elements of NLM's system reinvention initiative.

## Personnel

The NLM staff and many of the Library's users were saddened by the death on August 16 of Gerald R. Garner. Mr. Garner was a Library Technician well-known for his many years of excellent service at the Circulation Desk in the Main Reading Room. He was awarded the NIH Award of Merit posthumously.

In November 1998, Carolina A. Schragger, M.D., received a Postdoctoral Intramural Research Training Award (IRTA) from the National Center for Biotechnology Information (NCBI). Dr. Schragger received her medical degree from the New Jersey Medical School, University of Medicine and Dentistry of New Jersey, Newark, NJ in 1992. While at NCBI, she will write and edit content at the NCBI Web site and research topics from the biomedical sciences that demonstrate information resources at the NCBI.

In December 1998, Donald P. Jenkins, Ph.D. joined the High Performance Computing and Communications Office, LHCNCBC as a Special Expert. Dr. Jenkins received his Ph.D. in anatomical sciences in 1975 from the University of Washington in Seattle. At LHCNCBC, he will develop a cranio-facial teaching tool with Visible Human data accessed through the Internet and sponsor a research effort to develop anatomical methods designed to eliminate the methodological artifacts in the current dataset.

In December 1998, Steven A. Sullivan, Ph.D. joined NCBI as a Postdoctoral IRTA. Dr. Sullivan received his Ph.D. in anatomy and cell biology from the University of Michigan in 1993. At NCBI, Dr. Sullivan will identify fossil sequences by exhaustive comparison of protein sequences from evolutionary distant organisms and perform a comprehensive analyses of complete genomes.

In January 1999, Ms. Nancy Roderer was appointed Technical Information Specialist within the Division of Library Operations (LO) to serve as the NLM Library Associate Program Coordinator and Library Operations Research Advisor. Ms. Roderer has a B.S. degree in mathematics and computer science, an M.L.S. in information retrieval, and has published numerous papers and reports. Prior to her appointment at the NLM, Ms. Roderer was the Director of the Cushing/Whitney Medical Library at Yale University.

In January 1999, Leigh A. Riley, Ph.D. was appointed as NIH Staff Scientist with NCBI. Dr. Riley received her Ph.D. in human genetics and cell biology from the University of Texas. Before joining NLM, Dr. Riley was a GenBank indexer and a member of the team that answers indexing questions and writes indexing guidelines. Dr. Riley will continue to participate in all phases of processing of direct submissions to GenBank, and in training new indexers

in the mechanics of processing GenBank records.

In January 1999, Patrick Durand, Ph.D. joined the staff of NCBI as a Visiting Fellow. Dr. Durand, a native of Belgium, received his Ph.D. in molecular biophysics from the University Paris VI in June 1998. At NCBI, he will develop improved visualization tools for sequence alignment data and content-based filters to incorporate additional biological knowledge about the aligned sequences.

In February 1999, Hagit Shatkay-Reshef, Ph.D. joined NCBI as a Postdoctoral IRTA. Dr. Shatkay-Reshef, a native of Israel, received her Ph.D. in computer science from Brown University. Dr. Shatkay's current interest is to develop expertise in medical informatics and to apply hidden Markov models and other statistical and machine learning methodologies in this area. At NCBI, Dr. Shatkay-Reshef will work on the general problem of pattern recognition in natural language text.

In February 1999, Fasika Aklilu, Ph.D. joined NCBI as a Visiting Fellow. Dr. Aklilu, a native of Ethiopia, received her Ph.D. in experimental medicine from McGill University. Dr. Aklilu will be trained in software development at NCBI to build improved browsing and retrieval tools for genome related on-line resources.

In March 1999, Deborah G. Katz joined LHCNCBC as a Nurse Consultant. Ms. Katz received her BSN from the University of Connecticut and her MS from Renneselear Polytechnical Institute. In November 1987, Ms. Katz was recruited by the Division of AIDS, National Institute of Allergy and Infectious Diseases to work on a nationwide HIV information program. In her position at LHCNCBC, Ms. Katz will serve as a subject matter expert in the areas of clinical trials, drug development, and consumer health information.

In March 1999, Peter Kuehl joined NCBI as a Predoctoral IRTA. Mr. Kuehl received his Bachelor's Degree in molecular biology from Lehigh University and is a MD/Ph.D. student at the University of Maryland in the Predoctoral Molecular and Cellular Biology Training Program. His research will focus on sequence variations.

In March 1999, Boris A. Kirioutine, Ph.D. joined the NCBI as a Fogarty Visiting Fellow. Dr. Kirioutine received his Ph.D. in physics and mathematics from Moscow State University, Moscow, Russia. principal direction of Dr. Kirioutine's research at NCBI will involve application of advanced mathematical tools to genome analysis.

In March 1999, Jie Chen, Ph.D. joined NCBI as a Fogarty Visiting Fellow. Dr. Chen received her Ph.D. in computational molecular biology at Shanghai

Institute of Biochemistry. Her research program at NCBI will focus on computational approaches to the protein-protein docking problem.

In March 1999, Wataru Fujibuchi, Ph.D. joined NCBI as a Fogarty Visiting Fellow. Dr. Fujibuchi received his Ph.D. in biophysics from Kyoto University in Japan. At NCBI, he will develop new analytical techniques for DNA sequence analysis

In March 1999, David Roland Walker, joined NCBI as a Predoctoral IRTA. Mr. Walker received his BA in biology from Emory University and is a doctoral student in the Biology Department at Johns Hopkins University. Mr. Walker's research project is an annotation of the genome of the yeast *Saccharomyces cerevisiae*.

In April 1999, Capt. James E. Knoben, a USPHS Commissioned Corps Officer, transferred from the Health Care Financing Administration to the Biomedical Information Services Branch, Division of Specialized Information Services, NLM. At NLM, Capt. Knoben is responsible for conducting projects and providing strategies to evaluate drug toxicity information in the NLM toxicology databases. Capt. Knoben has a Doctor of Pharmacy degree from the University of California, and Masters of Public Health degree from Yale University.

In April 1999, Kathleen Cravedi, was selected for the newly established position of Public Liaison Officer within the Office of Communications and Public Liaison. Ms. Cravedi received her Masters degree from American University. As a Special Expert for the past three years, Ms. Cravedi has made significant contributions to NLM's public information and outreach-related functions. In her new position, Ms. Cravedi serves as the principal staff advisor to the NLM Director and senior staff on policy matters relating to communications plans involving the media.

In May 1999, Ms. Betsy L. Humphreys was appointed to the Senior Executive Service position of Associate Director for Library Operations, Division of Library Operations. Ms. Humphreys, who came to the NLM in 1973, is a Phi Beta Kappa graduate of Smith College and has a Masters of Library Science degree from the University of Maryland. She served as Deputy Associate Director for Library Operations since 1984 and held the concurrent position of Assistant NLM Director for Health Services Research Information since 1993.

In May 1999, Donna Maglott, Ph.D. joined NCBI as a Staff Scientist. Dr. Maglott received her Ph.D. in 1971 in zoology from the University of Michigan. Dr. Maglott has extensive experience in applying molecular biology to the development of gene

and genome-based information systems. At NCBI, she will work with the nomenclature groups of genome centers and the Human Nomenclature Committee.

In June 1999, Patricia Bosma was selected as Head, Selection and Acquisition Section, Technical Services Division, Division of Library Operations. Prior to her selection, Ms. Bosma served as Assistant Head of the Section for more than 18 years. Ms. Bosma is responsible for planning and directing the selection, acquisition, and processing of modern domestic and international monographic materials for the Library's collection. She received her MLS from the University of Maryland.

In June 1999, Dianne McCutcheon was selected as Head, Serial Records Section, Technical Services Division, Division of Library Operations. In her previous position as Assistant Head of the Serial Records Section for seven years, Ms. McCutcheon participated in the planning and management of selection, acquisition, and processing of over 23,000 biomedical serials publications. More recently, Ms. McCutcheon led the team that implemented NLM's new integrated library system (ILS). Ms. McCutcheon received her Masters degree from the School of Library Science, University of Illinois.

In June 1999, Andrew S. Trotman joined the staff of the Information Engineering Branch of NCBI as a Staff Scientist. Mr. Trotman, a native of New Zealand, received his Masters in computer science in 1992 from the University of Otago in New Zealand. Mr. Trotman will be working on a wide range of retrieval problems including making XML extensions to NCBI's Pub Med and Entrez retrieval systems.

In July 1999, Steven J. Phillips, M.D. joined the NLM as a Special Expert to serve as the principal medical advisor on research and medical education priorities. Dr. Phillips, who is board certified in general and thoracic surgery, served as Medical Director and Director of Research for the Iowa Heart Institute, Des Moines, Iowa. As Medical Director at Mercy Hospital in Des Moines, Dr. Phillips established the hospital's cardiac surgery program. His team pioneered techniques for emergency coronary bypass surgery as a treatment for evolving heart attacks, implanted the first artificial heart in Iowa, and invented the technology for percutaneous cardiopulmonary bypass.

In July 1999, Alexey S. Kondrashov, Ph.D. joined NCBI as a Senior Investigator. Dr. Kondrashov, a native of Russia, received his Ph.D. in evolutionary genetics from the Research Computer Center, Pushchino, and Institute of Developmental Biology, Moscow, in 1984. Dr. Kondrashov is an internationally known scientist in population genetics. His research has

had a particular focus on deleterious mutations and their consequence for evolutionary dynamics. Dr. Kondrashov's expertise will assist NCBI in building the NIH polymorphism database—dbSNP.

In July 1999, Julian D. Owens was appointed to the position of Special Expert, Division of Specialized Information Services. Mr. Owens received his Masters in public health from the University of Michigan and for the past several years he has served as a project leader with Andersen Consulting. As Special Expert, Mr. Owens will work on the design and implementation of outreach programs to assist the improvement of health of underserved populations.

In July 1999, Stephen W. Edwards, Ph.D. joined the staff of NCBI as a Postdoctoral IRTA. Dr. Edwards received his Ph.D. in pharmacology in 1999 from Vanderbilt University, Nashville, Tennessee. Dr. Edwards will work in the area of pharmacogenomics, studying the relationships between polymorphic sequence variation in genes encoding drug targets or metabolizing enzymes, and drug response phenotypes.

In August 1999, Maria Korab-Laskowska, Ph.D. joined NCBI as a Staff Scientist. Dr. Korab-Laskowska, a native of Poland, received her Ph.D. in natural science in 1983 from the Medical Academy in Warsaw, Poland. At NCBI, she will be working on database and software tool problems involving sequence and genetic information.

In August 1999, Ms. Susan Anderson joined the NLM for a two-year appointment under the Presidential Management Intern (PMI) Program. The PMI Program is designed to attract to the federal service outstanding graduate students from a wide variety of academic disciplines. Ms. Anderson received her M.S.L.S. from Catholic University. She has worked as a law librarian and as an intern at the D.C. Community Archives and has a particular interest in the preservation of nonprint materials.

In August 1999, Anthony F. Pirrone was selected for the position of Administrative Officer for NCBI. Prior to joining NCBI Mr. Pirrone worked as Administrative Officer with the National Institute of Diabetes and Digestive and Kidney Diseases, NIH.

In August 1999, Lukas Wagner, Ph.D. joined NCBI as a Staff Scientist. Dr. Wagner received his Ph.D. in physics from Ohio State University in 1996. At NCBI, Dr. Wagner is responsible for UniGene, a system for automatically classifying and analyzing transcribed sequences from human, mouse, and rat.

In August 1999, Francis A. Harrington, Ph.D. joined NCBI as a Postdoctoral IRTA. Dr. Harrington received her Ph.D. in mycology from Cornell University in 1996. At NCBI, her research will include the analysis of the

taxonomy of species in GenBank as well as the advancing of systematics of fungi.

In August 1999, Ron Edgar, Ph.D. joined NCBI as a Fogarty Visiting Fellow. Dr. Edgar, a native of Israel, received his Ph.D. degree from the Weizmann Institute of Science in 1998. At the NCBI, Dr. Edgar will conduct research on the storage, processing, analysis and dissemination of information derived from high-throughput gene expression technologies.

In September 1999, Susan Buyer was appointed to the permanent position of Chief, Office of Planning and Analysis, located within the Office of Health Information Programs Development. Over the past 4 years, she served as Acting Chief, Office of Planning and Analysis. Formerly, Ms. Buyer served as Deputy Assistant Director for Planning and Evaluation.

In September 1999, Colleen T. Webb joined NCBI as a Predoctoral IRTA. Ms. Webb received her M.S. in biology from the University of Oregon, Eugene, Oregon in 1994 and is a Ph.D. candidate in ecology and evolutionary biology at Cornell University. Ms. Webb will work on mathematical modeling to investigate the role of adaptive evolution in extinction and a research program in population genetics.

In September 1999, Paul A. Thiessen, Ph.D. joined NCBI as a Postdoctoral IRTA. Dr. Thiessen received his Ph.D. in chemistry from the University of Illinois in 1997. Dr. Thiessen will join the structure group at NCBI where he will work on improving the 3D structure visualization capability of NCBI's Entrez retrieval system.

In September 1999, Can Zhu Yang, Ph.D. joined NCBI as a Fogarty Visiting Fellow. Dr. Yang, a native of China, received her Ph.D. in bioinformatics in July 1999 from Yunnan University, People's Republic of China. At NCBI, Dr. Yang will be doing research and development work on the UniGene database.

In September 1999, Sarah Jo Wheelan, joined NCBI as a Pre-Doctoral IRTA. Ms. Wheelan received a BA in biochemistry, molecular, cellular and developmental biology and mathematics from the University of Colorado at Boulder in May 1995 and is a fourth year M.D.-Ph.D. student in the Human Genetics Program at The Johns Hopkins University School of Medicine. Her rotational projects at NCBI include comparative studies of cross-species orthologs, a positional cloning project, techniques to reconcile genetic mapping data and predicting the domain architecture of proteins a priori.

In September 1999, Deanna M. Church, Ph.D. joined the staff of the Information Engineering Branch of NCBI as a Postdoctoral IRTA. Dr. Church received her Ph.D. in biological sciences from the University of

California, Irvine in June 1997. While at NCBI, Dr. Church will focus on computational analysis of the human and mouse genomes.

In July 1999, Michael D. Moore, left NLM for a position with the Department of Commerce, Patent and Trademark Office. Mr. Moore served as Chief, Biomedical Files Implementation Branch, Division of Specialized Information Services, where he was responsible for the design, implementation and maintenance of the electronic databases used for the improvement and dissemination of toxicological and environmental health information.

#### *Retirements*

In December 1998, Lois Ann Colaianni retired from the federal government and her position as Associate Director for Library Operations, the position she held since 1984. Mrs. Colaianni first came to the NLM in January 1981 as Deputy Associate Director for Library Operations after serving as the Director of Libraries for Cedars-Sinai Medical Center in Los Angeles. She became Acting Associate Director in 1982 and was named Associate Director two years later. Mrs. Colaianni was responsible for planning, evaluating, implementing, and coordinating the program activities of the Division of Library Operations, including the formulation and establishment of policies, the development, analysis, and review of programs, the implementation of operating plans, the assessment of resource requirements, and the coordination of Library Operations programs with other components of NLM and with national major research libraries, national networks, and information centers in the U.S. and abroad.

In March 1999, Capt. Michael Weisberg retired from the USPHS Commissioned Corps with over 26 years of service. During his tenure at NLM, Dr. Weisberg managed the Learning Center for Interactive Technology and conducted work involving the application of computer-based learning in the health professions. Dr. Weisberg regularly hosted visitors from around the world, and designed and implemented several workshops and seminars on Internet-based learning.

In March 1999, William Willmering, Head of the Serial Records Section, Technical Services Division, Division of Library Operations, retired with over 18 years of service with the NLM. In addition to planning, controlling, and coordinating the activities of the Section, Mr. Willmering was responsible for a comprehensive program for the selection, acquisition, and processing of biomedical serial publications in all

languages in support of NLM's basic service programs.

In June 1999, Harold M. Schoolman, M.D., Deputy Director for Research and Education, retired from the Senior Executive Service with over 42 years of service in the Federal government. In July 1970, Dr. Schoolman joined the NLM to serve as an advisor and staff assistant in the area of program development and analysis directed toward insuring professional medical competence throughout the many components of the Library's programs. He provided continuing professional, technical, and medical evaluation of the program responsibilities of the Library. Dr. Schoolman was an active participant in Reinvention activities, and made major contributions to NLM, NIH, and DHHS policies regarding intellectual property issues.

In September 1999, Melvin L. Spann, Ph.D., Associate Director for the Division of Specialized Information Services, NLM, retired with 35 years of service from the Federal government. Dr. Spann began his career with NLM in 1976 as a Supervisory Chemist and was later promoted to the position of Associate Director in 1995. Dr. Spann was instrumental in the improvement of the exchange and dissemination of toxicological and environmental health information activities on a national and international scale. His involvement with the development of the Toxicology Information Outreach Program for Historically Black Colleges and Universities provided training for students and health professionals serving minority populations in the use of environmental and toxicological information resources. Dr. Stephen Phillips, Special Expert, is serving as Acting Director, SIS.

In September 1999, Frances H. Howard, Special Assistant to the Associate Director for Extramural Programs, retired from the Federal government with over 41 years of service. Ms. Howard joined the Library in February 1970. She was an effective spokesperson for NLM throughout her career and a highly effective liaison with the biomedical community, other Federal assistance programs and Congress.

#### *Awards*

The Annual Award for Outstanding Contributions in Health from the New York Academy of Medicine was presented to Dr. Donald A.B. Lindberg for his pioneering work in the application of computer technology to meet the needs of health care and biomedical research.

The Hammer Award from the Vice President of the United States was awarded to the National Library of Medicine. The award recognizes the Library for major improvements in its information services,

including making its popular MEDLINE database of journal article references and abstracts free and easier for the general public to use.

The Secretary's Award for Distinguished Service was presented to Ms. Marjorie A. Cahn (LO) and Ms. Maureen F. Prettyman (LHNCBC) as members of a National Guideline Clearinghouse Project Team for outstanding teamwork in the conceptualization, design, and development of the National Guideline Clearinghouse, world wide web-based repository for evidence-based clinical practice guidelines sponsored through a public/private partnership.

The NLM Board of Regents Award for Scholarship or Technical Achievement was awarded to Dr. Eugene V. Koonin (NCBI) in recognition of creative approaches to large scale analysis of bacterial genomes and to classification of protein families.

The Frank B. Rogers Award recognizes employees who have made significant contributions to the Library's fundamental operational programs and services. The recipient of the 1999 award was Ms. Eve Marie LaCroix (LO) for her rapid and highly successful design and implementation of MEDLINE*plus* which has significantly improved the public's access to high quality health information.

The NIH Director's Award was presented to Mr. Donald C. Poppke (OA) in recognition of outstanding administrative management skill and leadership in the National Library of Medicine.

The NLM Director's Award, presented in recognition of exceptional contributions to the NLM mission, was awarded to two employees: Ms. Susan P. Buyer (OD) in recognition of her leadership and extremely effective contributions to the National Library of Medicine's planning process and Mr. Theodore E. Youwer (OA) in recognition of his expert direction of the Office of Administrative Management Services and the consistently high quality service provided to all National Library of Medicine programs.

The NIH Merit Award was presented to seven employees: Mr. Joseph W. Hutchins (OCCS) for his commitment and accomplishments in replacing NLM's legacy systems with state-of-the-art commercial and custom designed client/server systems; Ms. Frances E. Johnson (EP) for consistent leadership and exceptional skill in administering NLM's grant programs for support of the nation's health science libraries; Mr. Gerald F. Garner (LO) for maintaining a consistent high level of service to the patrons in the NLM Reading Room; Ms. Barbara D. Shaw (LO) for a sustained high production level for the acquisition and processing of monographic orders and receipts for the NLM modern collection; and Dr. Elizabeth J. Van Lenten (LO) for

long-term contributions to the development and maintenance of the MEDLINE indexer workforce through training, contracting, and quality review. Ms. Carol Clausen (LO) and Ms. Linda G. Eisenstadt (OA) received the award as members of a group in support of the mission of the Office of Research Services NIH, in recognition of the development of a new cafeteria services program.

The NIH Champion of Diversity Award was presented to Mr. Donald C. Poppke (OA) for his active leadership role in promoting a successful diversity program within NLM and NIH.

The NIH Quality of Work Life Award was presented to the following LO team: Ms. Perlita M. Liwanag, Ms. Kristine M. Scannell, Ms. Frances A. Truong, and Ms. Monique C. Young in recognition of their outstanding efforts in planning, developing and guiding the implementation of the NLM Director's Employee Education Fund.

**TABLE 11**  
**Staff—FY 1999**

<b>Program</b>	<b>Full-Time Equivalents</b>
Office of the Director.....	19
Office of Public Information.....	8
Office of Administration.....	55
Office of Computer and Communications Systems.....	62
Extramural Programs.....	16
Lister Hill National Center for Biomedical Communications.....	76
National Center for Biotechnology Information.....	60
Specialized Information Services.....	31
Library Operations.....	274
<b>TOTAL FTEs.....</b>	<b>601</b>
<b>NLM Diversity Council</b>	

The NLM Diversity Council began the year by welcoming its new members. Cassandra Allen, Vivian Auld, Redmond Barnes, Dan Higgins, Sally Mooney, Margarita Ortiz and John Seachrist joined the Council for terms of two years beginning in January 1999 through December 2000. Other members who make up the Council are Mike Bumbray, Perlita Liwanag, Julia Royall and Francis Truong. The Council continues to receive support from its ex-officio members, Donald Poppke, David Nash and Nadgy Roey. Cassandra Allen continued in the role of Council chair and Julia Royall accepted Council Vice-Chair responsibilities.

## *FY 99 Accomplishments*

Coordinated NLM Director's Employee Education Fund activities. The Fund was established in November 1998. In its first year, the Fund supported 28 staff in taking 37 classes. Staff who have taken advantage of the Fund represent five different areas of NLM: 46 percent from the Division of Library operations, 27 percent from the Office of the Director, 16 percent from the National Center for Biotechnology Information, 8 percent from Specialized Information Services, and 3 percent from the Lister Hill National Center for Biomedical Communications. Undergraduate classes made up 70 percent of the classes supported. Staff participation is gradually increasing and the Diversity Council will continue efforts to publicize the availability of the Fund.

Provided briefings on the activities of NLM's Diversity Council or related issues to:

- Vice President Gore's National Partnership for Reinventing Government
- Diversity Task Force
- NIH Diversity Council
- Calvin Coolidge High School Adopt-a-School Program summer students
- NIH Ombudsman

Continued monthly brown bag sessions for NLM staff through May 1999. Sessions included:

- Jumpstarting Your Career
- Creating an Individual Development Plan
- Writing a Resume
- Managing job Stress
- Overview of Mentoring

Arranged for a briefing to the Diversity Council and the NLM Senior Staff from Cheryl Kelley of Vice President Gore's National Partnership for Reinventing Government, Diversity Task Force.

Provided Diversity Training to the Council. Since only two of the current Council members had received the initial diversity training in October 1996, the entire Council was provided with training to ensure that all members understood the concepts of diversity.

Continued to promote additional staff recipes to NLM Cooks!, a listing on the Diversity Council's Intranet Homepage.

Continued to coordinate scheduling of CPR training classes for NLM staff.

Initiated the development of additional projects that will carry over into FY 2000:

- NLM After Hours: a Website on NLM's Intranet that features the creative talents of NLM's staff. NLM After Hours is a tool to help staff better know and appreciate one another.
- Getting to Know NLM: a monthly lecture series designed to make staff more knowledgeable about the work performed in the different operational units at NLM and the skills needed to perform this work. Ultimately, the Council hopes to promote diversity while broadening the staff's knowledge of NLM.
- Mentoring: An investigation into the pros and cons of pursuing a mentoring program at the NLM.
- Deaf Awareness: an invitation has been extended to Dr. I. King Jordan, President of Gallaudet University in Washington, D.C. to speak at the NLM.

## APPENDIX 1: REGIONAL MEDICAL LIBRARIES

- 1. MIDDLE ATLANTIC REGION**  
The New York Academy of Medicine  
1216 Fifth Avenue  
New York, NY 10029-5283  
(212) 822-7396 FAX (212) 534-7042  
States served: DE, NJ, NY, PA  
*URL: <http://www.nlm.nih.gov/mar>*
- 2. SOUTHEASTERN/ATLANTIC REGION**  
University of Maryland at Baltimore  
Health Science and Human Services Library  
601 Lombard Street  
Baltimore, MD 21201-1583  
(410) 706-2855 FAX (410) 706-0099  
States served: AL, FL, GA, MD, MS, NC,  
SC, TN, VA, WV, DC, VI, PR  
*URL: <http://www.nlm.nih.gov/sar>*
- 3. GREATER MIDWEST REGION**  
University of Illinois at Chicago  
Library of the Health Sciences (M/C 763)  
1750 West Polk Street  
Chicago, IL 60612-7223  
(312) 996-2464 FAX (312) 996-2226  
States served: IA, IL, IN, KY, MI, MN,  
ND, OH, SD, WI  
*URL: <http://www.nlm.nih.gov/gmr>*
- 4. MIDCONTINENTAL REGION**  
University of Nebraska Medical Center  
Leon S. McGoogan Library of Medicine  
Regional Medical Library  
986706 Nebraska Medical Center  
Omaha, NE 68198-6706  
(402) 559-4326 FAX (402) 559-5482  
States served: CO, KS, MO, NE, UT, WY  
*URL: <http://www.nlm.nih.gov/mr>*
- 5. SOUTH CENTRAL REGION**  
Houston Academy of Medicine-  
Texas Medical Center Library  
1133 M.D. Anderson Boulevard  
Houston, TX 77030-2809  
(713) 799-7880 FAX (713) 790-7030  
States served: AR, LA, NM, OK, TX  
*URL: <http://www.nlm.nih.gov/scr>*
- 6. PACIFIC NORTHWEST REGION**  
University of Washington  
Regional Medical Library, HSLIC  
Box 357155  
Seattle, WA 98195-7155  
(206) 543-8262 FAX (206) 543-2469  
States served: AK, ID, MT, OR, WA  
*URL: <http://www.nlm.nih.gov/pnr>*
- 7. PACIFIC SOUTHWEST REGION**  
University of California, Los Angeles  
Louise M. Darling Biomedical Library  
Box 951798  
Los Angeles, CA 90025-1798  
(310) 825-1200 FAX (310) 825-5389  
States served: AZ, CA, HI, NV and U.S.  
Territories in the Pacific Basin  
*URL: <http://www.nlm.nih.gov/psr>*
- 8. NEW ENGLAND REGION**  
University of Connecticut Health Center  
Lyman Maynard Stowe Library  
263 Farmington Avenue  
Farmington, CT 06030-5370  
(860) 679-4500 FAX (860) 679-1305  
States served: CT, MA, ME, NH, RI, VT  
*URL: <http://www.nlm.nih.gov/ner>*

## APPENDIX 2: BOARD OF REGENTS

The NLM Board of Regents meets three times a year to consider Library issues and make recommendations to the Secretary of Health and Human Services affecting the Library

### Appointed Members:

BOND, Enriqueta, Ph.D. (Chair)  
President  
Burroughs Wellcome Fund  
Durham, NC

BARUCH, Jordan, Sc.D.  
President, Jordan Baruch Associates  
Washington, D.C.

BUNTING, Alison, M.L.S.  
Associate University Library for Science  
Louise Darling Biomedical Library  
University of California, Los Angeles  
Los Angeles, CA

FONSECA, Raymond J., D.M.D.  
Department of Oral and Maxillofacial Surgery  
Univ. of Pennsylvania School of Dental Medicine  
Philadelphia, PA

FOSTER, Henry, M.D., Ph.D.  
Senior Advisor to the President on Teen & Youth  
Issues  
Department of Health and Human Services  
Washington, D.C.

GAGE, John  
Director, Science Office  
Sun Microsystems Computer Corporation  
Palo Alto, CA

KLEIN FEDYSHI, Michele, MSLS  
Falk Library of the Health Sciences  
University of Pittsburgh  
Pittsburgh, PA

LEDERBERG, Joshua, Ph.D.  
Sackler Foundation Scholar  
Rockefeller University  
New York, NY

NEWHOUSE, Joseph BA, Ph.D.  
Director, Division of Health Policy and Research  
Harvard University  
Boston, MA

PARDES, Herbert, M.D.  
Vice President for Health Sciences  
Dean of Faculty of Medicine  
College of Physicians and Surgeons  
Columbia University  
New York, NY 10032

### Ex Officio Members:

Librarian of Congress

Surgeon General  
Public Health Service

Surgeon General  
Department of the Air Force

Surgeon General  
Department of the Navy

Surgeon General  
Department of the Army

Under Secretary for Health  
Department of Veterans Affairs

Assistant Director for Biological Sciences  
National Science Foundation

Director  
National Agricultural Library

Dean  
Uniformed Services University of the Health Sciences

### **APPENDIX 3: BOARD OF SCIENTIFIC COUNSELORS/ LISTER HILL CENTER**

The Board of Scientific Counselors meets periodically to review and make recommendations on the Library's intramural research and development programs.

**Members:**

JAFFE, Conrade C., M.D.  
Director, Center for Advanced Instructional Media  
Yale University School of Medicine  
New Haven, CT

MARSHALL, Joanne G. Ph.D.  
Dean, School of Information & Library Science  
University of North Carolina  
Chapel Hill, NC

MASYS, Daniel R., M.D.  
Director of Biomedical Informatics  
School of Medicine  
University of California at San Diego  
La Jolla, CA

MITRA, Sunanda, Ph.D.  
Professor of Electrical Engineering  
Texas Tech University  
Lubbock, TX

SIEVERT, MaryEllen C., Ph.D.  
Professor of Library and Information Science  
University of Missouri  
Columbia, MO

SRINIVASAN, Padmini, Ph.D.  
School of Library & Information Science  
University of Iowa  
Iowa City, IA

## **APPENDIX 4: BOARD OF SCIENTIFIC COUNSELORS/ NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION**

The National Center for Biotechnology Information Board of Scientific Counselors meets periodically to review and make recommendations on the Library's biotechnology-related programs.

### **Members:**

ROBERTS, Richard J., Ph.D. (Chair)  
Research Director  
New England Biolabs  
Beverly, MA

BUETOW, Kenneth H., Ph.D.  
Chief, Laboratory of Population Genetics  
National Cancer Institute  
Bethesda, MD

DELISI, Charles, Ph.D.  
Dean, College of Engineering  
Boston University  
Boston, MA

LEE, Christopher J., Ph.D.  
Assistant Professor  
Molecular Biology Institute  
University of California Los Angeles  
Los Angeles, CA

MATISE, Tara Cox, Ph.D.  
Assistant Professor  
Laboratory of Statistical Genetics  
The Rockefeller University  
New York, NY

## APPENDIX 5: BIOMEDICAL LIBRARY REVIEW COMMITTEE

The Biomedical Library Review Committee meets three times a year to review applications for grants under the Medical Library Assistance Act.

### Members:

BASLER, Thomas G., Ph.D. (Chair)  
Chair, Department of Library Science & Informatics  
Medical University of South Carolina  
Charleston, SC

ASH, Joan S., Ph.D.  
Associate Professor  
Library and Medical Informatics  
Oregon Health Sciences University  
Portland, OR

CHUEH, Henry C., M.D.  
Co-Director, Laboratory of Computer Science  
Assistant Professor of Medicine  
Harvard Medical School  
Boston, MA

CHUTE, Christopher G., Dr.P.H., M.D.  
Section Head and Professor  
Medical Informatics  
Mayo Foundation  
Rochester, MN

DALRYMPLE, Prudence, Ph.D.  
Dean and Associate Professor  
Graduate School of Library Information Science  
Dominican University  
River Forest, IL

DIMITROFF, Alexandra, Ph.D.  
Associate Professor  
School of Library Science  
University of Wisconsin  
Milwaukee, WI

FLORANCE, Valerie, Ph.D.  
Project Director  
Association of American Medical Colleges  
Washington, DC

FUCHS, Rainer T., Ph.D.  
Vice President and CIO

ARIAD Pharmaceuticals  
Hoechst-ARIAD Genomics Center  
Cambridge, MA

HOLST, Ruth  
Director, Library Services  
Columbia Hospital  
Milwaukee, WI

HUANG, H.K., D.Sc.  
Director, Radiological Informatics  
University of California at San Francisco  
San Francisco, CA

LATTMAN, Eaton E., Ph.D.  
Professor and Chair  
Department of Biophysics  
Johns Hopkins University  
Baltimore, MD

MCGOWAN, Julie J., Ph.D.  
Director, Ruth Lilly Medical Library  
Indiana University School of Medicine  
Indianapolis, IN

MILLER, Randolph A., M.D.  
Chairman, Division of Biomedical Informatics  
Vanderbilt University Medical Center  
Nashville, TN

MOULT, John, Ph.D.  
Professor  
Center for Advanced Research in Biotechnology  
Rockville, MD

NILAND, Joyce C., Ph.D.  
Chair, Division of Information Sciences  
City of Hope National Medical Center  
Duarte, CA

OHNO-MACHADO, Lucila, M.D., Ph.D.  
Assistant Professor, Radiology Department  
Brigham and Women's Hospital  
Harvard University  
Boston, MA

### Members:

ORTHNER, Helmuth , Ph.D.  
Professor, Department of Health Informatics  
University of Alabama  
Birmingham, AL

PINSKY, Seth, Ph.D.  
Senior Director  
Merck and Company, Inc.  
Rahway, NJ

RINDFLEISCH, Thomas  
Director, Lane Medical Library  
Stanford University  
Stanford, CA

SAHNI, Sartaj K., Ph.D.  
Distinguished Professor  
Computer & Information Science  
University of Florida  
Gainesville, FL

TANG, Paul C., M.D.  
Medical Director, Clinical Informatics  
Palo Alto Medical Foundation  
Los Altos, CA

## APPENDIX 6: LITERATURE SELECTION TECHNICAL REVIEW COMMITTEE

The Literature Selection Technical Review Committee meets three times a year to select journals for indexing in *Index Medicus* and MEDLINE.

### Members:

CLEVER, Linda Hawes, M.D. (Chair)  
Chair, Dept. of Occupational Health  
California Pacific Medical Center  
San Francisco, CA 94120

BIRKMEYER, John D., M.D.  
Assistant Professor of Surgery  
Veterans Affairs Medical Center  
White River Junction, VT

CABELLO, Felipe C., M.D.  
Dept. of Microbiology and Immunology  
New York Medical College  
New York, NY

COLLEN, Morris F., M.D.  
Consultant and Director Emeritus  
Kaiser Permanente Medical Care Program  
Oakland, CA

COOPER, James N., M.D.  
Director, INOVA Institute of Research  
Chairman, Department of Medicine  
Fairfax Hospital  
Falls Church, VA

COPELAND, Robert L., Ph.D.  
Associate Professor of Pharmacology  
Howard University School of Medicine  
Washington, D.C.

EPSTEIN, Neal, M.D.  
Cardiology Branch  
National Heart, Lung, and Blood Institute  
Bethesda, MD

FUNK, Mark E.  
Samuel J. Wood Library  
Weill Medical College  
Cornell University  
New York, NY

LI, Yihong, Ph.D.  
Assistant Professor  
Oral Biology Department  
University of Alabama School of Dentistry  
Birmingham, AL

O'DONNELL, Anne Elizabeth, M.D.  
Assistant Professor  
Pulmonary and Critical Care Medicine  
Georgetown University School of Medicine  
Washington, D.C.

PICOT, Sandra J. Fulton, Ph.D.  
Associate Professor  
School of Nursing  
University of Maryland  
Baltimore, MD

WILLIAMS, Benjamin T., M.D.  
Professor of Information Science  
University of Illinois  
Champaign, IL

## APPENDIX 7: EXTRAMURAL PROGRAMS GRANTS—FY 1999

Note: Awards preceded by an \* are first year, or competing awards; all others are continuation of awards made in prior years. The amount of award represents Total Costs (Direct Costs and Facilities and Administrative Costs).

### **Integrated Advanced Information Management Systems (IAIMS) Grants, Phase I**

Burton, Terrance, \$146,711  
5 G08 LM06277-02  
Health Sciences Center IAIMS Program  
West Virginia University, Morgantown, WV

Panko, Walter B., \$150,000  
5 G08 LM06307-02  
IAIMS Planning At UIC  
University of Illinois, Chicago, IL

\*Brennan, Patricia F., \$137,812  
1 G08 LM06647-01A1  
Health Science IAIMS Initiative  
University of Wisconsin, Madison, WI

\*Byrd, Gary D., \$138,024  
1 G08 LM06763-01  
Integrated Health Information For Western NY  
Suny at Buffalo, Buffalo, NY

\*Berns, Kenneth, \$149,508  
1 G08 LM06821-01  
University of Florida IAIMS Planning Grant  
University of Florida, Gainesville, FL

\*Wilson, H. D., \$149,998  
1 G08 LM06824-01  
N.D. IAIMS Health Information Infrastructure  
University of North Dakota, Grand Forks, ND

### **Integrated Advanced Information Management Systems (IAIMS) Grants, Phase II**

Mitchell, Joyce A., \$550,000  
5 G08 LM05415-06  
Missouri IAIMS  
University of Missouri, Columbia, MO

Stead, William W., \$550,000  
5 G08 LM05443-07  
Fast Track Provision of IAIMS  
Vanderbilt University, Nashville, TN

Paton, John A., \$542,068  
5 G08 LM05583-05  
Yale University school of Medicine  
New haven, CT

Niland, Joyce C., \$550,000  
5 G08 LM05697-04  
City of Hope National Medical Center  
Duarte, CA

Friedman, Charles P., \$550,000  
5 G08 LM06625-02  
Advancing The Academic Mission  
University of Pittsburgh, Pittsburgh, PA  
Resource Information Systems Grants

Dennis, Sharon E., \$137,783  
5 G08 LM05684-05  
Model Multimedia Support Center  
University of Utah, Salt Lake City, UT

\*Jones, Michael O., \$119,944  
2 G08 LM06218-04  
Automated Archive of American Traditional Med.  
University of California, Los Angeles

Mueller, Christine A., \$138,962  
5 G08 LM06239-03  
Health Information — Rural Nurse Practitioners  
University of Minnesota, Minneapolis

Cohen, Betty K., \$61,014  
5 G08 LM06575-02  
New Haven Public Health Database  
Yale University, New Haven, CT

Lemkau, Henry L., \$110,486  
5 G08 LM06583-02  
"POINTIS"  
University of Miami, Miami, FL

Leipzig, Rosanne, \$110,024  
5 G08 LM06648-02  
Evidence Based Medicine Infor. & Resource Ctr.  
New York Academy of Medicine, New York

Mutka, Martin, \$82950  
 5 G08 LM06697-02  
 Oncall II  
 SUNY at Buffalo, Buffalo, NY

\*Boilard, David W., \$118,403  
 1 G08 LM06765-01  
 Health Information Network For NW Ohio  
 Medical College of Ohio, Toledo

\*Cassell, K, \$150,000  
 1 G08 LM06799-01  
 Community Health Onsite Information Centers  
 The New York Public Library, New York, NY  
 Resource Information Access Grants

\*Binderman, Mary S., \$11,952  
 1 G07 LM06580-01A1  
 On-Line Access to the Archives of Occ. Therapy  
 American Occupational Therapy Foundation  
 Bethesda, MD

\*Richwine, Margaret, \$197,312  
 1 G07 LM06611-01A1  
 Shared Hosp. Electronic Library of Southern Ind.  
 Indiana University, Indianapolis, IN

\*Auflick, Patricia A., \$48,000  
 1 G07 LM06616-01A1  
 Pinal County Information Infrastructure Dev.  
 University of Arizona, Tucson

\*Carney, John, \$35,956  
 1 G08 LM06618-01A1  
 Access To Medical Literature Project  
 Charles Drew University, Los Angeles, CA

\*Ryle, Pat, \$11,980  
 1 G07 LM06622-01A1  
 Expanding Access To On-Line Resources  
 Greater Staples Hospital, Staples, MN

\*Beyer, Kirk, \$12,000  
 1 G07 LM06652-01A1  
 P. C. Lakota Campus NLM Connectivity Project  
 Presentation College, Aberdeen, SD

\*Smith, Cheryl, \$119,691  
 1 G07 LM06815-01  
 Triple A (Access, Ability, and Assistance)  
 Southwest Louisiana Area AHEC  
 Lafayette, LA

Internet Connection Grants

\*Whitman, Nancy I., \$48,500  
 1 G07 LM06740-01  
 Lynchburg Neighbornet  
 Lynchburg College, Lynchburg, VA

Wilson, Stephen S., \$11,450  
 3 G08 LM06760-01S1  
 Integration of a Research LAN and the Internet  
 La Jolla Institute For Allergy/Immunology  
 San Diego, CA

\*Bear, Louise J., \$29,831  
 1 G08 LM06788-01  
 Sir's Rural Health Clinics  
 Lassen Indian Health Clinic, Susanville, CA

\*Schwegler, Raymond W., \$29,940  
 1 G08 LM06803-01  
 Manor Dental Clinic  
 Manor Junior College, Jenkintown, PA

\*Sasmor, Jeannette L., \$27,894  
 1 G08 LM06805-01  
 Yavapai Comm. Health Information Network  
 Yavapai College, Prescott, AZ

\*Cleaves, Clayton T., \$29,166  
 1 G08 LM06818-01  
 Maine Indian Health Center  
 Pleasant Point Health Center, Perry, ME

\*Harvell, Jonathan E., \$49,968  
 1 G08 LM06820-01  
 Internet Initiative For West Central Maine  
 Franklin Memorial Hospital, Farmington, ME

\*Genova, Peter L., \$29,722  
 1 G08 LM06844-01  
 Campus Wide Internet Connection  
 Long Beach Med. Center, Long Beach, NY

\*Rousey, James, \$22,585  
 1 G08 LM06855-01  
 Internet Connection  
 Madison Co. Health Dept., Richmond, KY

\*Stephenson, Michael, \$49,885  
 1 G08 LM06857-01  
 NE Michigan Rural Health Services  
 Thunder Bay Com. Health Ser., Hillman, MI

\*Mcnamee, James E., \$30,000  
 1 G08 LM06969-01  
 Rural Medical Education  
 University of South Carolina, Columbia

\*Aiu, Pauaallaokani, \$28,030  
 1 G08 LM06989-01  
 Papa Ola Lokahi Internet Connection  
 Papa Ola Lokahi, Honolulu, HI

\*Murrin, Patricia, \$30,000  
 1 G08 LM06971-01  
 EMS Internet Connection Project  
 Emergency Med. Services, San Diego, CA

\*Deleel, Phillip G., \$48,457  
 1 G08 LM06972-01  
 North Country Hospitals Connection Project  
 Canton-Potsdam Hospital, Potsdam, NY

\*Abbott, Patricia, \$48,509  
 1 G08 LM06977-01  
 The Governors Well Mobile  
 University of Maryland, Baltimore

\*Stephens, Tim, \$44,100  
 1 G08 LM06980-01  
 T1 Line For Anson And Beaufort Counties  
 University of North Carolina, Chapel Hill

\*Lane, Kevin, \$49,939  
 1 G08 LM06981-01  
 Cook County Hosp.-Rush U. Internet Project  
 Hektoen Institute/CCH, Chicago, IL

\*Anderson, Anita, \$29,611  
 1 G08 LM06982-01  
 Internet Access Extension  
 Valley Health System, Inc., Huntington, WV

\*Rodriguez, Patricia, \$28,420  
 1 G08 LM06995-01  
 Rappahannock AHEC, Montross, VA

\*Bernstein, Alan, \$22,359  
 1 G08 LM07000-01  
 Health Department ME & LAB  
 Monroe County Health Department, Rochester, NY  
 Publication Grants  
 Kiple, Kenneth, \$20,000  
 5 R01 LM06574-02  
 History and Culture of Food and Nutrition  
 Bowling Green State U., Bowling Green, OH

Speaker, Susan, \$35,000  
 5 R01 LM06590-02  
 Talk About Drugs —Policy 1875-1990  
 Middleburg, PA

\*Reynolds, P. P., \$73,102  
 1 R01 LM06617-01A1  
 Racial Integration of Medicine in the U.S.  
 Johns Hopkins University, Baltimore, MD

\*Benedict, Susan C., \$35,973  
 1 R01 LM06653-01A1  
 Nurses and the Nazi Euthanasia Programs  
 Medical University of South Carolina, Charleston

\*Schmanhann, Jeremy D., \$59,950  
 1 R01 LM06708-01  
 3-D MRI Atlas of the Human Cerebellum  
 Massachusetts General Hospital, Boston  
 \$59,850

**Medical Informatics Research Grants**

Jaffe, Conrade, \$244,421  
 5 R01 LM05007-10  
 Indexing Image Databases  
 Yale University, New Haven, CT

Miller, Webb, \$414,231  
 5 R01 LM05110-11  
 Software For Analyzing Biosequence Data  
 Pennsylvania State University, University Park

Berner, Eta S., \$125,137  
 5 R01 LM05125-08  
 Physician Use of Diagnostic Decision Support System  
 Data  
 University of Alabama, Birmingham

Jelliffe, Roger W., \$513,220  
 5 R01 LM05401-08  
 Decision Supports and Databases for Drug Dosage  
 University of Southern California, Los Angeles

\*Selker, Harry, \$407,380  
 2 R01 LM05607-04A1  
 Assessing Mathematical Models for Medical Events  
 New England Medical Center Hospital, Boston, MA

Toga, Arthur, \$240,849  
 5 R01 LM05639-05  
 Digital Representation & Visualization of Human Brain

University of California, Los Angeles

Musen, Mark A., \$546,864  
5 R01 LM05708-05  
Software Architecture for Guideline Directed Therapy  
Stanford University, Stanford, CA

Brutlag, Douglas L., \$56,010  
3 R01 LM05716-05S1  
Multiple Representations of Biological Sequences  
Stanford University, Stanford, CA

Goldbaum, Michael, \$96,832  
3 R01 LM05759-09S1  
Structured Analysis of the Retina  
University of California, La Jolla

Marshall, Bryan, \$235,385  
5 R01 LM05997-03  
Information Integration & Virtual Therapy in the SICU  
University of Pennsylvania, Philadelphia

Miller, Randolph A., \$496,898  
5 R01 LM06226-03  
Patient Care Provider Order Entry with Tactical  
Support  
Vanderbilt University, Nashville, TN

Good, Walter F., \$150,972  
5 R01 LM06236-05  
Non ROC Measures for Evaluating Image Compression  
University of Pittsburgh, Pittsburgh, PA

\*Bowden, Douglas, \$16,000  
2 R01 LM06243-04  
Spatial-Symbolic Brain Info. Mgmt. System  
University of Washington, Seattle

Altman, Russ, \$330,540  
5 R01 LM06244-04  
Representing Biological Data for Molecular Modeling  
Research Management Group, Stanford, CA

Brennan, Patricia, \$407,646  
5 R01 LM06249-03  
Computer Support - Home Care of Patients  
University of Wisconsin, Madison

Huang, H. K., \$258,460  
5 R01 LM06270-03  
Digital Hand Atlas in Assessment of Skeletal Dev.  
University of California, San Francisco

Friedman, Carol, \$209,130  
5 R01 LM06274-03  
Unlocking Data from Med. Records with Text  
Processing  
Research Foundation – SUNY, New York, NY

Hersh, William, \$186,284  
5 R01 LM06311-03  
New Model and Approach to Retrieval System Eval.  
Oregon Health Sciences University, Portland

Brinkley, James F., \$296,940  
5 R01 LM06316-03  
Structure Based Visual Access To Biomedical Info.  
University of Washington, Seattle

Sonnenberg, Frank, \$205,403  
5 R01 LM06321-02  
Decision Analytic Support for Clinical Guidelines  
UMDNJ/RWJ Medical School, Piscataway, NJ

Grobe, Susan J., \$164,512  
5 R01 LM06325-03  
Automated Analysis of Intervention Narrative  
University of Texas, Austin

Luther, Paul, \$130,250  
5 R01 LM06326-03  
Three Dimensional Reconstruction of Synapses  
University of Maryland, Baltimore

Goodwin, Linda, \$308,857  
5 R01 LM06488-03  
Medical Perinatal Knowledge Building  
Duke University, Durham, NC

Gustafson, David, \$382,980  
5 R01 LM06533-02  
Platform Impact-Breast Cancer  
University of Wisconsin, Madison

Ohno-Machado, Lucila, \$249,197  
5 R01 LM06538-02  
Component Based Tools—Connectionist Class.  
Brigham And Women's Hospital, Boston, MA

Haug, Peter, \$163,209  
5 R01 LM06539-03  
Semantic Parser for Medical Free Text  
LDS Hospital, Salt Lake City, UT

Brodley, Carla, \$329,999  
5 R01 LM06543-02

Content Based Image Retrieval for Med. Databases  
Purdue Research Foundation, West Lafayette, IN

Kohane, Isaac S., \$302,068  
5 R01 LM06587-02  
Health Info. Identification & De-Identification Toolkit  
Children's Hospital, Boston, MA

Gardner, Reed, \$487,463  
5 R01 LM06591-02  
Clinical Software Quality Review Process  
LDS Hospital, Salt Lake City, UT

Mc Keown, Kathleen, \$311,274  
5 R01 LM06593-02  
Multimedia Summaries of PAT STAT POSTBYPA  
Columbia University, New York, NY

\*Shortliffe, Edward, \$650,000  
1 R01 LM06594-01  
A Shared Internet Server For Delivering Guidelines  
Stanford University, Stanford, CA

Metaxas, Dimitri, \$219,923  
5 R01 LM06638-02  
Analysis of #30 Cardiac Motion from Tagged MRI  
University of Pennsylvania, Philadelphia

\*Pustejovsky, James, \$343,110  
1 R01 LM06649-01A1  
Automated Knowledge Extraction for Biomedical Lit.  
Brandeis University, Waltham, MA

Miller, Perry, \$308,075  
5 R01 LM06682-02  
Support for Computer Based Clinical Guidelines  
Yale University, New Haven, CT

Cooper, Gregory F., \$193,887  
5 R01 LM06696-02  
Cause and Effect from Clinical Data  
University of Pittsburgh, Pittsburgh, PA

Ezquerro, Norberto F., \$341,417  
5 R01 LM06726-02  
Distributed Cardiac Image Bases  
Georgia Tech Research Corp, Atlanta, GA

Neuwald, Andrew, \$334,876  
5 R01 LM06747-02  
Sequence Based Prediction of Protein Function  
Cold Spring Harbor Laboratory, NY

\*Buchanan, Bruce G., \$208,234  
1 R01 LM06759-01  
Data Mining and Model Building  
University of Pittsburgh, Pittsburgh, PA

\*Sim, Ida, \$171,444  
1 R01 LM06780-01  
Electronic Knowledge Based —Randomized Trials  
University of California, San Francisco

\*Rosse, Cornelius, \$359,190  
1 R01 LM06822-01  
Foundation Model of Anatomy  
University of Washington, Seattle

\*Salzberg, Steven L., \$267,416  
1 R01 LM06845-01  
Computational Modeling of Genes and Gene Structure  
The Institute For Genomic Research, Rockville, MD

\*Mcalindon, Timothy E., \$170,382  
1 R01 LM06856-01  
Feasibility of Internet-Based Trials for Osteoarthritis  
Boston University, Boston, MA

\*Panniers, Teresa, \$210,479  
1 R01 LM06967-01  
Decision Support for Neonatal Intensive Care Nursing  
New York University, New York, NY  
Medical Informatics First Awards

Shiffman, Richard N., \$113,371  
5 R29 LM05552-03  
Knowledge Processing for Clinical Practice Guidelines  
Yale University, New Haven, CT

Gorman, Paul N., \$110,318  
5 R29 LM05663-05  
Information Seeking in Primary Care  
Oregon Health Sciences University, Portland, OR

Yang, Yiming, \$103,033  
5 R29 LM05714-05  
LLSF Mapping for Indexing & Retrieval of Medicine  
Carnegie Mellon University, Pittsburgh, PA

Goldstein, Richard A., \$93,545  
5 R29 LM05770-05  
Protein Structure Prediction  
University Of Michigan, Ann Arbor

Johnson, Stephen B., \$121,326  
5 R29 LM05783-05

Medical Information through Natural Language  
Columbia University, New York, NY

Tong, David A., \$148,802  
5 R29 LM06004-05  
Interpretation of Intracardiac Electrograms  
Southwest Research Institute, San Antonio, TX

Wagner, Michael M., \$100,760  
5 R29 LM06233-03  
Belief Network Based Reminder Systems  
University of Pittsburgh, Pittsburgh

Shahar, Yuval, \$116,424  
5 R29 LM06245-04  
Knowledge Based Abstraction of Clinical Data  
Stanford University, Stanford, CA

Wong, Stephen T., \$92,083  
5 R29 LM06300-02  
Multimodality Neuroimaging Database System  
University of California, San Francisco

Stanley, Mary, \$94,224  
5 R29 LM06322-03  
3D Dynamic Anatomy - A Virtual Reality Prototype  
University Of Central Florida, Orlando, FL

Brummer, Marijn, \$109,045  
5 R29 LM06486-03  
4D Visualization of Congenital Heart Disease  
Emory University, Atlanta, GA

#### **Biotechnology Resource Grants**

Herzenberg, Leonore, \$324,801  
5 R01 LM04836-09  
Knowledge Base Support for Flow Cytometry  
Stanford University, Stanford, CA

Ledley, Robert S., \$1,226,259  
5 P41 LM05798-05  
Protein Information Resource  
National Biomedical Research Foundation  
Washington, DC

\*Markley, John, \$410,001  
2 P41 LM05799-04  
Biological Magnetic Resonance Data Bank  
University of Wisconsin, Madison

Roberts, Richard J., \$160,773  
5 P41 LM05800-05

Rebase-The Restriction Enzyme Database  
New England Biolabs, Beverly, MA

\*Pagon, Roberta A., \$244,908  
2 P41 LM06001-04  
HELIX: Genetic Testing Resource  
Children's Hospital And Medical Center  
Seattle, WA

Pagon, Roberta A., \$608,526  
5 P41 LMO6029-03  
Genline—Electronic Clinical Genetics Knowledge Base  
University of Washington, Seattle

#### **Conference Grant**

\*Dunker, A. K., \$25,993  
1 R13 LM06766-01  
Pacific Symposium On Biocomputing 1999-2001  
Washington State University, Pullman

#### **Research Training In Medical Informatics**

Shortliffe, Edward, \$739,172  
5 T15 LM07033-16  
Graduate Training In Medical Information Sciences  
Stanford university, Stanford, CA

Gatewood, Lael. \$494,800  
5 T15 LM07041-16  
Research Training in Medical Informatics  
University of Minnesota, Minneapolis

Miller, Perry, \$359,229  
5 T15 LM07056-13  
Medical Informatics Research Training at Yale  
Yale University, New Haven, CT

Friedman, Charles, \$734,232  
5 T15 LM07059-13  
Pittsburgh Medical Informatics Training Program  
University of Pittsburgh, Pittsburgh

Downs, Stephen, \$593,966  
5 T15 LM07071-08  
Duke-UNC Training Program In Medical Informatics  
University of North Carolina, Chapel Hill

Cimino, James, \$828,880  
5 T15 LM07079-07  
Medical Informatics Research Training  
Columbia University, New York, NY

Spackman, Kent A., \$373,337; \$149,756 –07S1  
3 T15 LM07088-07S1  
5 T15 LM07088-08  
Training Program In Health Informatics  
Oregon Health Sciences University, Portland, OR

Mitchell, Joyce, \$398,356  
5 T15 LM07089-08  
Medical Informatics Research Training Program  
University of Missouri, Columbia

Greenes, Robert, \$1,419,659  
5 T15 LM07092-08  
Harvard-MIT-NEMC Res. Trng. in Health Informatics  
Harvard Medical School, Boston, MA

Gorry, G. Anthony, \$496,976  
5 T15 LM07093-08  
Training Program in Computational Biology & Med.  
William Marsh Rice University, Houston, TX

McDonald, Clement, \$240,163; \$69,777 –02S1  
3 T15 LM07117-02S1  
5 T15 LM07117-03  
Regenstrief Med. Informatics Research Fellowships  
Indiana University, Indianapolis

Gardner, Reed, \$688,040  
5 T15 LM07124-03  
University of Utah Medical Informatics Training  
University of Utah, Salt Lake City

#### **Applied Medical Informatics Fellowships**

Steinberger, Eileen K., \$53,694  
5 F38 LM00061-02  
Expert System — Lymphoma and Leukemia  
University of Maryland, Baltimore

Lehmann, Christopher, \$46,700  
5 F38 LM00064-02  
A Web-Based Case Simulation  
Johns Hopkins University, Baltimore, MD

Schubart, Jane R., \$34,046  
5 F38 LM00068-02  
Evaluation of an Info. Resource for Improving Care  
University of Virginia, Charlottesville

Terrazas, Enrique, \$55,395  
5 F38 LM00072-02  
Web Based Laboratory Order Entry System

University Of California, San Francisco, CA

\*Moore, Carlton R., \$76,710  
1 F38 LM00078-01  
Database Of Free Medications For Indigent Patients  
Mount Sinai School Of Medicine, New York, NY

\*Hawkins, Fanny, \$65,500  
1 F38 LM00080-01A1  
Ambulatory Medical Record Strategy  
Baylor College Of Medicine, Houston, TX

\*Gobar, Glenna M., \$31,673  
1 F38 LM00084-01  
National Companion Animal Health Surveillance Sys.  
University Of California, Davis

\*Zhou, Jie, \$53,214  
1 F38 LM00085-01  
A Novel Medical Knowledge Retrieval System  
University of California, Davis, CA

\*Lober, William B., \$65,000  
1 F38 LM00086-01  
Knowledge Based Image Annotation  
University Of Washington, Seattle

#### **Small Business Innovative Research (SBIR)**

Baclawski, Kenneth, \$10,000  
3 R43 LM06665-02S1  
Biomedical Science Information Retrieval & Mgmt.  
Jarg Corporation, Lincoln, MA

\*McKay, Colin, \$99,189  
1 R43 LM06777-01  
Case-In-Point Psychiatric Nursing Educ. Multimedia  
Smilex, Inc., Houston, TX

\*Posse, Christian, \$100,000  
Propurview: Data Analysis Using Projection Pursuit  
Talaria, Inc., Seattle, WA

\*Chen, Richard O., \$201,889  
1 R43 LM06878-01  
Genenet-A Knowledge Base of Functional Genetic Data  
Ingenuity Systems, Inc., Redwood City, CA

\*Whitaker, Virginia, \$118,931  
1 R43 LM07016-01  
Computer-Aided Assessment Of Early Vocabulary  
Trefoil Corp., Orono, ME

## Digital Library 2 Grants of Interest to Biomedicine

### *Biomedical Domains:*

Gorman P., Oregon Health Sciences University; Tracking Footprints Through Information Space; \$250,000

Wiederhold, G., Stanford; Trusted Image Dissemination; \$200,000

McKeown, C., Columbia; A Patient Care Digital Library; \$1,000,000

Schatz, B., U. Illinois; Building a Medical Interspace Test bed; \$600,000

Rowe, T., U. Texas; Comparative Vertebrate Morphology [uses Visible Human]; \$125,000

Chen, H., U. Arizona; Classification of Medline; \$125,000

Buneman, P., U. of Penn; Data Provenance [of biotechnology data]; \$125,000

### *Non-Biomedical Domains, But Topics Applicable To Biomedicine:*

Lagoze, C., Cornell; Security and Reliability in Digital Libraries; \$900,000

Huang, T., U. Illinois; Video Structuring and Searching at the Semantic Level; \$200,000

Palakal, M., Indiana U.; Information Filtering System for Digital Libraries; \$100,000

### *Mixed Domains:*

Wilensky, R., U. California, Berkeley-, [multiple domains, including Neuroscience]; \$1,100,000

## National Heart Attack Alert Initiative

### *Phase I - Contracts Awarded*

RFP: BAA/RFP NLM 99-108/VMS

- Johns Hopkins Medical Institutions and the Johns Hopkins University Applied Physics Laboratory
- The MEDSTAT Group
- Mitretek Systems
- TRW, Inc.
- University of Pennsylvania

### *Phase II - Contracts Awarded*

The eight contractors who had received awards in FY 1998 were allowed to compete for a Phase II award to continue the work. Only two were funded by the end of the 1999 fiscal year because of budget constraints, but others in this group may receive funding in FY 2000.

Barnett, G. O.

1N01LM93535

Informatics for the National Heart Attack Alert Program

Massachusetts General Hospital  
Boston, MA

Cimino, James

3N01LM93534

Informatics for the National Heart Attack Alert Program

Columbia University Health Sciences  
New York, NY

# INDEX

## A

ACTIS .....	24
African American History Month.....	17
Agency for Health Care Policy and Research ...	17, 55
Agency for Toxic Substances and Disease Registry .....	22, 25
AHEC.....	15
AIDS Community Outreach Program .....	24
AIDS Information Services .....	24
AIDSDRUGS.....	24
AIDSTRIALS .....	24
Alternatives to Animal Testing .....	25
American Dental Association.....	13
American Hospital Association.....	13
American Library Association .....	16
AsnLib.....	49
Associate Fellowship.....	17
Asthma exhibit .....	16, 33
ATIS.....	24
Audiovisual Program Development Branch.....	26
Avery.....	31
AVLINE.....	11
Awards .....	74

## B

BankIt.....	48
Basic research .....	49
Bibliographic Control .....	8
BIOETHICSLINE.....	54
Biomedical Ethics .....	54
Biomedical Information Science Technology Initiative .....	53
Biomedical Library Review Committee.....	55
Biotechnology Informatics.....	53
BISTI .....	53
BLAST.....	45, 47
Board of Regents.....	1, 55, 59
Board of Scientific Counselors .....	26, 50
Boston University Medical Center.....	15
Breath of Life.....	16, 33

## C

Cancer Genome Anatomy Project.....	44, 46
Cataloging.....	9
CATLINE .....	11

CCRIS.....	23
Chemical Structure Server .....	24
ChemID.....	24
CHEMID.....	22
ChemIDplus .....	24
Citation Data Creation and Maintenance System....	63
Classics of Traditional Chinese Medicine.....	16
Clinical Elective in Medical Informatics.....	38
Clinical trials.....	29
Cn3D.....	48
Coffee Break Web page .....	45
Cognitive Science Branch .....	26
COGs .....	47
Collection Development.....	6
Communications Engineering Branch.....	26
Computer Center .....	58
Computer Science Branch.....	27
Conference Grants.....	54
Consumer groups .....	16
Consumer health informatics.....	29
Current Bibliographies in Medicine.....	13
Customer Services Section.....	14
CustomerQ.....	14

## D

DART.....	23
dbEST .....	44
dbSNP .....	45
dbSTS .....	44
Dental journals .....	10
Department of Health and Human Services .....	17
Department of Veterans Affairs .....	24
Desktop Services Section.....	67
Digital Libraries Initiative.....	54
Digital library research.....	31
Digitized x-rays.....	36
DIRLINE.....	23
Distributed Learning Environments .....	33
Diversity Council .....	75
DOCLINE.....	14, 59, 64
Document Delivery .....	13
Document Imaging Laboratory .....	40
DOCUSER.....	59, 64
DocView .....	2, 32

## E

Electronic publications.....	6
ELHILL.....	11, 61, 69
Elizabeth Blackwell .....	16
EMIC .....	23
Entrez.....	44, 46, 48, 69
Environmental Protection Agency .....	23, 25
Environmental Teratology Information Center .....	23
EtherMed .....	33
European Molecular Biology Laboratory.....	43
Exhibits .....	16, 50
Expressed Sequence Tags .....	44
Extramural bioinformatics.....	51
Extramural Programs Division.....	52

## F

FIRST award.....	56
Frankenstein: Discovering the Secrets of Nature ...	34
Frequently Asked Questions .....	14
Friends of the NLM.....	16
Fruit fly .....	44

## G

Gateway .....	11, 30
GenBank .....	43, 48, 50
Gene Express .....	46
Gene Map.....	45
GeneMap.....	45
GENE-TOX .....	23
Genome Survey Sequences .....	44
Georgetown University .....	54
Ghana .....	2
Grateful Med.....	11
Guideline Clearinghouse .....	30

## H

Hammer Award.....	59, 74
Hazardous Substances Data Bank .....	22
Health Data Standards Committee .....	17
Health Hotlines .....	23
Health Informatics.....	17
health services research.....	17
HealthStar .....	69
HealthSTAR.....	11
HelpDesk.....	67
High Performance Computing and Communications .....	18, 27, 41
High-Throughput Genomic sequences .....	44
Historical collections.....	7
History of Medicine Division.....	16

Hospital and Health Administration Index.....	13
Houston Academy of Medicine.....	15
HSTAT.....	11, 30, 60
Human Genome Project .....	44
Human Genome Resources Web page .....	45

## I

IAIMS grants.....	52
Idaho State University.....	15
Image Processing Laboratory.....	40
Image Storage and Transmission Optimization.....	35
Images from the History of Medicine.....	9
Impromptu.....	62
Index Medicus.....	13
Index to Dental Literature .....	13
Indexing .....	10
Indexing Initiative .....	11, 29
Indian Health Service.....	15
Informatics research.....	53
Information Access Grants.....	52
Information Systems Laboratory .....	67
Institute of Medicine .....	22, 24
Integrated Advanced Information Management Systems.....	52
Integrated Library System.....	59, 60, 69
International DOCLINE Libraries.....	2
International Loansome Doc .....	62
Internet Connection Grants .....	52
Internet connectivity.....	2, 15
Internet Grateful Med.....	11, 14, 24, 29
Inventory Control.....	62
IRIS .....	23

## J

Japan .....	43
-------------	----

## K

Kellogg Foundation.....	16
Kenya .....	1
Knowledge Source Server.....	28

## L

Learning Center for Interactive Technology .....	32
Lederberg.....	31
Legacy systems .....	63
LinkOut .....	44, 46
List of Journals Indexed in Index Medicus .....	13
List of Serials Indexed for Online Users .....	13
Lister Hill National Center for Biomedical Communications.....	26

Literature Search Technical Review Committee	59
Literature Selection Technical Review Committee	10
Loansome Doc	2, 14, 62
LOCATORplus	11, 14, 60, 69
LocusLink	45
Long Range Plan	1
Lower Mississippi Delta	25

## M

Malaria	1
Malaria Genetics	47
Malaria Research and Training Center	1
Mali	1
MARC	12
MARS Production Facility	40
Medical Article Record System	31
Medical College of Ohio	15
Medical Imaging Tools	37
Medical Informatics	53
Medical Informatics Training	38
Medical Library Assistance Act	52
Medical Library Association	16, 17
Medical Subject Headings	8, 63, 69
MEDLINE	2, 11, 29, 30, 31, 44, 48, 59
MEDLINEplus	6, 7, 12, 16, 24, 62
Mesh2000	63
MetaMap	29
Metathesaurus	9, 28
Michigan State University	15
Minorities	53
Minority students	38
Montana State University	15
Multilateral Initiative on Malaria	1

## N

National Academy of Sciences	41
National AIDS Minority Information and Education Program	25
National Cancer Institute	23, 46
National Center for Biotechnology Information	43
National Center for Toxicological Research	23
National Health and Nutrition Examination Survey	36
National Heart Attack Alert Program	55
National Heart, Lung, and Blood Institute	16, 55
National Information Center on Health Services Research	30
National Institute of Allergy and Infectious Diseases	16, 47
National Institute of Environmental Health Sciences	16, 23
National Institute of Occupational Safety and Health	23

National Library of Medicine Classification	9
National Medical Librarians Month	17
National Network of Libraries of Medicine	3, 13, 14
National Online Training Center	17
National Performance Review	60, 69
National Research Council	41
National Science Foundation	54
Native Americans	4
Next Generation Internet	38, 41
NHANES	36
NICHSR	15, 17
NIDCR	41
NLM Classification	62
NLM web site	61
NLM Web site	13, 63

## O

Oak Ridge Associated Universities	50
Object Oriented Information Retrieval Project	30
Office of Computer and Communications Systems	58
Office of High Performance Computing and Communications	27
OMIM	45
Online indexing system	10
Online Public Access Catalog	60
Onsite users	13
Oracle DBMS	63
Ordering/Inventory Control System	62
Outreach Activities	3, 23
Overnight Photocopy Service	60

## P

Pacific Northwest	4
Pacific Northwest Regional Medical Library	4
Partners in Information Access	15, 24
Portable Core Library	49
Preservation	8
Profiles in Science	10, 31
public libraries	16
Public libraries	4
Public Services Division	62
Publication Grant Program	54
PubMed	2, 10, 11, 14, 17, 46, 69
PubMed Central	6, 10, 46

## R

Reading room	6
Reference and Customer Service	14
RefSeq	45
Regional Medical Libraries	15
Reinvention	11, 12, 60

Relais .....	14, 60, 63
Resource Grants .....	52
Retirements .....	73
Rodbell.....	31
RTECS .....	23

**S**

SAGEmap .....	46
Science .....	45
Semantic Knowledge Representation.....	28
Sequence Tagged Site .....	44
Sequin .....	49
SERHOLD .....	59, 64
SERLINE .....	11
SilverPlatter.....	14
Smart card technology.....	30
Space needs.....	6
SPECIALIST .....	28
Specialized Information Services.....	15, 22
Superfund.....	23
SUPERLIST.....	22
System Reinvention.....	29, 69
System Security.....	38
Systems Reinvention .....	59
Systems Technology Branch .....	64

**T**

Taxonomy database.....	48
Telemedicine.....	18, 33, 40
Telepathology .....	33
Telephone service .....	14
Texas Woman's University .....	15
TLC.....	32
Toxic Chemical Release Inventory .....	23
Toxicology and Environmental Health Information Program.....	22
Toxicology Data Network.....	24
Toxicology Information Outreach Project .....	25

TOXICOLOGY TUTOR.....	25
TOXLINE .....	22
TOXLINE65 .....	22
TOXNET .....	23, 60
Training.....	17, 24, 27, 52
Training in Informatics.....	52
Training of Minorities.....	53
Trans-Pacific Digital Libraries Experiment .....	39
TRI.....	23

**U**

UMLS .....	63
UMLS Knowledge Sources.....	12
Unified Medical Language System .....	8, 27, 54, 63
UniGene .....	45
University of Connecticut .....	15
User Support .....	51

**V**

vBNS.....	39
Visible Human Atlas of the Head and Neck.....	42
Visible Human Project .....	33, 34, 55
Visitors' Program.....	50
Voyager Integrated Library System .. 5, 13, 59, 60, 63	

**W**

WebMIRS .....	36
Woman's History Month.....	17

**Y**

Year 2000 Concerns.....	58, 59, 65, 68
-------------------------	----------------

**Z**

Zimbabwe.....	2
---------------	---